



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**EFFECTS OF THE WIRELESS CHANNEL, SIGNAL
COMPRESSION AND NETWORK ARCHITECTURE ON
SPEECH QUALITY IN VOIP NETWORKS**

by

Tiantioukas Nikolaos

June 2007

Thesis Advisor:

Murali Tummala

Thesis Co-Advisor:

John C. McEachen

Approved for public release, distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 2007	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE: Effects of the Wireless Channel, Signal Compression and Network Architecture on Speech Quality in Voip Networks			5. FUNDING NUMBERS	
6. AUTHOR: Tantioukas Nikolaos				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) Voice over Internet Protocol (VoIP) telephony is an emerging technology slowly finding its way into military applications. It provides several advantages over PSTN but comes short on performance, quality of service and availability. The purpose of this thesis is to measure the quality of voice in VoIP communications. More specifically it investigates the effects of wireless channel conditions as well as channel coding and compression on the received speech quality. Both simulation and experimentation are conducted using Matlab code and Speex software and across commercial VoIP networks. Simulation shows that fading channel parameters can heavily affect the quality of received speech. Speech compression results in bit rate gain, but, on the other hand, the signal becomes more sensitive to errors. The performance of an outdoor wireless network is better than that of an indoor network. The VoIP network architecture can affect the received speech quality on a long-distance connection.				
14. SUBJECT TERMS Voice Quality, Digital Communications, Channel Coding, VoIP, Fading Channel, Voice Compression			15. NUMBER OF PAGES 105	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release, distribution is unlimited

**EFFECTS OF THE WIRELESS CHANNEL, SIGNAL COMPRESSION AND
NETWORK ARCHITECTURE ON SPEECH QUALITY IN VOIP NETWORKS**

Tiantioukas Nikolaos
Lieutenant, Hellenic Navy
B.A., Hellenic Naval Academy, 1997

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

**NAVAL POSTGRADUATE SCHOOL
June 2007**

Author: Tiantioukas Nikolaos

Approved by: Murali Tummala
Thesis Advisor

John C. McEachen
Co-advisor

Jeffrey B. Knorr
Chairman, Department of Electrical and Computer Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Voice over Internet Protocol (VoIP) telephony is an emerging technology slowly finding its way into military applications. It provides several advantages over PSTN but comes short on performance, quality of service and availability.

The purpose of this thesis is to measure the quality of voice in VoIP communications. More specifically it investigates the effects of wireless channel conditions as well as channel coding and compression on the received speech quality. Both simulation and experimentation are conducted using Matlab code and Speex software and across commercial VoIP networks.

Simulation shows that fading channel parameters can heavily affect the quality of received speech. Speech compression results in bit rate gain, but, on the other hand, the signal becomes more sensitive to errors. The performance of an outdoor wireless network is better than that of an indoor network. The VoIP network architecture can affect the received speech quality on a long-distance connection.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	 THESIS OBJECTIVE	2
B.	 RELATED WORK	2
C.	 THESIS ORGANIZATION.....	3
II.	VOICE OVER INTERNET PROTOCOL	5
A.	 VOIP OVERVIEW	5
1.	Protocols.....	10
a.	<i>RTP</i>	<i>11</i>
b.	<i>RUDP.....</i>	<i>12</i>
2.	Signaling in VoIP	13
a.	<i>H.323</i>	<i>14</i>
b.	<i>SIP</i>	<i>14</i>
B.	 VOICE CODING	17
1.	Speech Compression	17
2.	Standards of Speech Compression	17
a.	<i>Linear Predictive Coding</i>	<i>17</i>
b.	<i>Code Excited Linear Prediction (CELP).....</i>	<i>19</i>
c.	<i>CELP Variations</i>	<i>20</i>
C.	 VOICE RECOGNITION	21
1.	Different Approaches and Constraints	21
2.	Dynamic Time Warping (DTW).....	22
3.	Hidden Markov Model (HMM)	22
D.	 QOS.....	23
1.	VoIP Availability.....	23
2.	VoIP and E.911.....	24
3.	Factors Affecting QoS.....	25
a.	<i>Packet Delay.....</i>	<i>25</i>
b.	<i>Packet Delay Variation</i>	<i>25</i>
c.	<i>Packet Loss.....</i>	<i>26</i>
d.	<i>Type of Codec</i>	<i>26</i>
4.	Clarity of Received Speech and Methods of Measuring It.....	26
E.	 SUMMARY	28
III.	WIRELESS NETWORKS.....	29
A.	 DIGITAL WIRELESS COMMUNICATION NETWORK	29
B.	 CHANNEL.....	30
1.	Attenuation and Noise	31
2.	Multipath	32
3.	Fading.....	33
C.	 MODULATION AND CHANNEL CODING	35
1.	Modulation.....	35
2.	Channel Coding.....	37

D.	WIRELESS STANDARDS OF INTEREST	38
1.	IEEE 802.11	38
2.	802.16.....	41
E.	SUMMARY	43
IV.	RESULTS	45
A.	MATLAB SIMULATION.....	46
1.	Rician Fading Channel in AWGN.....	47
2.	Rician Fading Channel in AWGN and Convolutional Coding.....	48
3.	Simulation Results	48
a.	<i>Rician Fading Channel: K factor.....</i>	<i>48</i>
b.	<i>Rician Fading Channel: SNR</i>	<i>49</i>
c.	<i>Rician Fading Channel With Convolutional Coding.....</i>	<i>50</i>
d.	<i>Secondary Path Delays in a Rician Fading Channel.....</i>	<i>52</i>
e.	<i>Effect of Secondary Path Signal Strength in a Rician Fading Channel</i>	<i>53</i>
f.	<i>Remaining Speech after Decompression as a Function of Compression Ratio</i>	<i>54</i>
g.	<i>Remaining Speech after Decompression as a function of Compression Ratio With and Without Channel Coding</i>	<i>55</i>
B.	VOIP EXPERIMENTS OVER A WIRELESS LAN	57
1.	Skype Service.....	57
2.	Vonage Service	60
C.	VOIP EXPERIMENT ON A WAN.....	63
D.	SUMMARY	65
V.	CONCLUSIONS	67
A.	SIGNIFICANT RESULTS.....	67
B.	FUTURE WORK.....	68
	APPENDIX A	71
A.	SPEEX OVERVIEW	71
B.	DRAGON NATURALLY SPEAKING	71
	APPENDIX B	73
	MATLAB CODE	73
	LIST OF REFERENCES	81
	INITIAL DISTRIBUTION LIST	85

LIST OF FIGURES

Figure 1.	Basic VoIP Communication System.....	6
Figure 2.	IP to IP VoIP Implementation Interconnecting two Different LANs with Internet.....	7
Figure 3.	IP to IP with Wireless or Satellite Implementation Included on the Interconnection of two Separate LANs.....	8
Figure 4.	VoIP to PSTN: Interconnection between a VoIP User and a Traditional Telephony User.....	9
Figure 5.	Telco to VoIP to Telco Implementation, Interconnecting two Users of traditional Telephony Using VoIP.....	10
Figure 6.	OSI Layer and VoIP-used protocols.....	11
Figure 7.	IP/UDP/RTP Header Format.....	12
Figure 8.	RUDP Header Format.....	13
Figure 9.	VoIP Signaling Procedure Including Call Setup and Voice Exchange Phase.....	14
Figure 10.	Complete Phone Application Use Including Look Up Phase and Call Setup Phase.....	15
Figure 11.	SIP Redirection from User Agent to Proxy Server, Redirect Server and Database Server.....	16
Figure 12.	Model for Human Speech Production Process used by Linear Predictive Coding.....	18
Figure 13.	Block Diagram of Federal Standard 1016 Using CELP and Featuring two Codebooks, Stochastic and Adaptive.....	20
Figure 14.	HMM with Six States Labeled as Integers.....	23
Figure 15.	Block Diagram of Digital Voice Communication System.....	29
Figure 16.	Implementation of a Wireless Network Using a Sum of Interconnected Nodes.....	30
Figure 17.	Schematic of Multipath Channel Where the Signal Follows a Direct, a Reflected and a Diffracted Path.....	33
Figure 18.	Constellation Diagram of BPSK, QPSK and 16QAM Modulation.....	37
Figure 19.	A $(n,k,K) = (2,1,3)$ Convolutional Encoder Implementation Using a Shift Register.....	38
Figure 20.	Schematic Diagram of IEEE 802.11a Basic Function Blocks.....	41
Figure 21.	Interconnection of the IEEE 802.11-based LANs using the 802.16 as a Point-to-point Link.....	42
Figure 22.	Overall Setup Used for Simulation/experiment Voice Transmission over a Packet Switched Wireless Communication System.....	46
Figure 23.	Matlab and Additional Simulation Software Setup used to Simulate VoIP over a Wireless Network.....	47
Figure 24.	The Bit Error Rate as a function of the Ratio of Dominant over Secondary Path (K factor) for the Rician Fading Channel based on 50 Monte Carlo Simulation Runs.....	49

Figure 25.	Effect of SNR on Bit Error Rate for a Rician Fading Channel based on 50 Monte Carlo Simulation Runs	50
Figure 26.	Effects of SNR on the BER for a Rician Fading Channel with Convolutional Coding, and Speech Coding based on 50 Monte Carlo Simulation Runs.....	52
Figure 27.	Effect of Secondary Path Delays on the BER of a Rician Fading Channel based on 50 Monte Carlo Simulation Runs	53
Figure 28.	BER as a Function of Secondary Path Gain for a Rician Fading Channel based on 50 Monte Carlo Simulation Runs	54
Figure 29.	Effect of Compression Ratio on the Remaining Speech after Decompression based on 50 Monte Carlo Simulation Runs	55
Figure 30.	Effect of Compression Ratio on the Remaining Speech after Decompression with Channel Coding based on 50 Monte Carlo Simulation Runs.....	56
Figure 31.	Skype Implementation with Two Users Interconnected with VoIP Inside the Same LAN.....	58
Figure 32.	Amount of Remaining Speech versus Average Receiver Signal Strength for Skype Measurement Setup based on 12 Monte Carlo Runs	59
Figure 33.	Vonage Implementation as an Alternative Platform to Skype. The Two Users are in the Same LAN but the Signal Travels through Vonage Servers in New Jersey during Both the Call Setup and the Data Exchange Phase	61
Figure 34.	Amount of Remaining Speech versus Average Receiver Signal Strength for Vonage Experimentation.....	62
Figure 35.	VoIP Measurements on a WAN Implementation. The Two Users are Located in Monterey, California and Athens, Greece, Respectively. One User Transmits a Recorded Message through VoIP Using Skype and the Other Records it.	63
Figure 36.	Percentage of Remaining Speech for a Speech Signal Transmitted in Greece and Recorded in USA as Shown in Figure 31	64
Figure 37.	Schematic Diagram of Speex and Dragon Naturally Speaking Interconnection with Matlab	72

LIST OF TABLES

Table 1.	Popular Codecs with Their Frame Size and Bit Rate	27
Table 2.	Data Rate and Frequencies Used in 802.11 Physical Layer	41

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Prof. Murali Tummala for his guidance and support throughout this thesis preparation. I would also like to thank Mr. Bob Broadston for his precious technical support and Prof. John McEachen for his comments and corrections. Next I would like to thank my fellow student Capt. David Manka for his continuous help and cooperation.

Finally, I would like to thank the inspiration of my life and dreams, my wife Marina, for her continuous support and patience, during our stay in Monterey, and especially during my thesis writing.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Voice communication has been continually evolving since Alexander Bell's discovery. For a long period of time, circuit switched networks dominated the transmission of voice. Circuit switched networks were also used as a medium for data transmission. The picture today is totally different, with packet switched networks supporting both data and voice communications.

An emerging technology that uses packet switched networks for voice transmission is Voice over Internet Protocol (VoIP) telephony. It is widely used in the commercial sector and is slowly finding its way into military applications. It is already being used on a trial basis on the battlefield as well as permanent installations. The advantages that make this technology preferable to traditional telephony are low cost, use of the existing infrastructure, and the ability to add new applications without additional cost. The combined use of VoIP and wireless networks provides further advantages since it provides seamless communication without the use of physical cabling among units, which enables faster deployment. On the other hand, the drawbacks of VoIP are the inferior performance and quality of service as well as limited availability when compared to traditional telephone networks. In order for VoIP to be able to compete with the traditional telephone networks, it must improve the quality of service and availability especially when used with wireless networks.

With the above as motivation, the objective of this thesis is to measure the quality of voice in VoIP communications. More specifically, an investigation of IP-based voice communication with emphasis on the effects of a wireless channel on the quality of the received speech is attempted, and the effects of voice signal compression and wireless channel conditions as well as channel coding on the voice quality and recognition are investigated through both simulation and experimentation. Simulation is implemented using Matlab and Speex software, and the experiments are conducted on commercial VoIP networks.

Simulation demonstrates that by increasing the SNR of the dominant path in a Rician fading channel, the BER of the transmission decreases. On the same channel, an increase in the secondary path delay variation causes an increase in the BER of the signal, and, as the signal strength of the secondary paths increases, the BER increases as well. There is no significant difference between the BER of a compressed and uncompressed speech signal when passing through the same channel; however, the amount of audible distortion is higher when the speech signal is compressed. Compressing a speech signal results in a gain in bit rate, but, on the other hand, the signal becomes more sensitive to errors. Next, the effect of compression ratio on the speech quality is examined, and simulation results show that as the signal is compressed at higher rates and passed through the same channel, quality of the received speech deteriorates. When channel coding is used, not only the speech quality has improved, but also the errors are eliminated in comparison to speech transmission without channel coding.

Experiments were conducted on two different platforms, namely Skype and Vonage, to investigate the effects of architectures of the two providers on the received speech quality and the effectiveness of VoIP during a 24-hour period on a long-distance connection. Experimental results indicate that performance of an outdoor wireless network is better than that of an indoor network due to the effect of multipath occurring indoors. Comparing the results for Skype and Vonage, it is noticed that Skype achieves a slightly better performance for both the outdoor and indoor environment. The architecture of the Vonage network causes additional delay, path loss and multiple hops that contribute to a higher packet loss. Experiments of VoIP over the Internet for a long distance communication indicate that speech quality follows a random pattern due to the dynamic nature of the Internet traffic. Degradation of speech quality is observed during the rush hours due to network congestion since during these rush hours the signal has to travel through slower lines causing additional delay and thus a decrease in speech quality.

This work is based on the need to investigate the effects of a wireless channel, speech compression and channel coding on the quality of the received speech.

Investigation is conducted through both simulation and experimentation. The results of these simulations and experiments were reported and the need for future work is discussed.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Voice over Internet Protocol (VoIP) Telephony is a technology already in use by the commercial sector. Following the commercial sector on this expanding technology, the armed forces have adopted this new digital communication concept in a limited capacity. This transformation took place using mainly existing infrastructure, meaning at a minimum cost for the required task. New applications are added without additional cost and without interrupting the flow of existing applications. [1].

The Marine Corps implemented VoIP in their deployments to provide integrated and seamless communications to all levels of command [2]. The main purpose of using the VoIP technology is to provide voice communication down to the last unit, combined with data exchange and without the need to deploy extra infrastructure wherever the unit is deployed. When a unit is deployed and interconnected with a data network to the main information grid, members of the unit can communicate with one another and also with any senior authority as needed. Furthermore, the extra bandwidth remaining beyond speech communication can be used to automatically send additional battlefield information in the form of video or data. Some examples of such information are temperature, level of supplies, and other relevant sensor data.

Even though deployment of VoIP in battle conditions is more spectacular and draws attention immediately, the beneficial contribution to permanent installations in places such as naval bases or airports must not be downplayed. It is easy to see why VoIP is being widely deployed if one takes into account the cost savings compared to traditional telephony, combined with the benefits of additional applications. The installation cost is also decreased since the existing network infrastructure is used and skills and manpower needed for administration are reduced [3].

The real evolution in military communications comes from the combined use of VoIP and wireless networks. A deployed unit does not need to carry the copper/fiber cable, all units are interconnected without physical cabling between them, and the Navy, Army and Air Force personnel can communicate seamlessly. The need for wireless is definitely more crucial for units that are deployed in foreign territories in an

expeditionary manner. There is no infrastructure available and even if there were any, it will most likely be destroyed during the initial takeover, making the need for ad hoc communications vital. VoIP over wireless is an effective solution for this scenario for two reasons. First, the man-hours and skills needed to deploy and effectively administrate the network are minimal, which is an important factor when the available human power is limited. Second, it allows for further network expansion when backups arrive and when the need for seamless communication between the two networks is immediate.

A. THESIS OBJECTIVE

Considering the need for further expansion of VoIP, especially through wireless networks, it is essential to investigate the determining factors between VoIP and traditional telephony. The factors that public switched telephone network (PSTN) shows superior performance compared to VoIP are the quality of service provided and the network availability. In order for VoIP to be able to compete with PSTN, these two factors must reach a level close to that of the PSTN [4].

Having that in mind, the focus of this thesis is to measure the comprehension and recognition of voice through digital communications. More specifically, IP-based voice communication is analyzed with emphasis on the effects of a wireless channel on quality of the received speech. Factors that affect the quality of the received voice, such as channel status, compression ratio, and channel coding, are quantified. The effects of voice signal compression and wireless channel conditions as well as channel coding on the voice quality and recognition are investigated through both simulation and experimentation.

B. RELATED WORK

During the last decade, speech quality in VoIP has been a rich research area. Some important work in the area is discussed below.

The effects of passive interruptions and communication delay on a phone conversation quality have been subject of investigation [5]. The results indicate that there is a strong relationship between the number of passive interruptions on the conversation

and the quality of the received speech. On the other hand, the delay induced on the conversation has small influence on the perceptual quality of the conversation. A further analysis of the factors affecting the voice quality of VoIP can be found in [6]. The factors analyzed are delay, jitter, packet loss, link errors, echo and Voice Activity Detection (VAD). Ways to smooth the negative effects of these factors are presented.

An investigation of retransmission schemes that can help recover corrupted packets is attempted with a focus on avoiding long retransmission delays in [7]. The results show that the retransmission performance depends on the quality of the link as well as the introduced delay. The transmission of VoIP packets in a concatenated manner is proposed in order to increase the throughput [8]. The proposed aggregation is achieved by transmitting multiple VoIP packets in a multicast packet so that the throughput of the VoIP implementation is increased.

The effects of voice transmission over secure wireless networks are investigated and the results show that security choices of a VoIP network can affect the VoIP design [9].

An evaluation of real time control protocol's (RTCP) effectiveness is attempted in [10], and the results show that even though RTCP is effective for low delay networks, it can be inaccurate for networks with large, volatile delays.

Similar to the above mentioned related work, this thesis investigates the speech quality in a VoIP network. In contrast to previous efforts, emphasis is given to the effects of wireless channel as well as the effects of signal compression and channel compression on speech quality. Furthermore, the effects of different VoIP network architectures are investigated. The work of [10] is further expanded with an experimental study on the RTCP effectiveness on networks with large propagation delays.

C. THESIS ORGANIZATION

The thesis is organized as follows. Chapter II introduces the Voice over Internet Protocol (VoIP). An introduction to transport protocols, signaling, and voice coding as well as voice recognition and quality of service are attempted. Chapter III describes the wireless networks and more specifically introduces the concept of a digital

communication network and focuses on the wireless channel and the factors that affect it. Next, modulation and channel coding are discussed and some wireless standards of interest are presented. Chapter IV presents the Matlab simulation setup and the simulation results derived from it. This is followed by experiments over commercial VoIP networks and the results so obtained. Chapter V summarizes the thesis and proposes future work. Appendix A includes a brief description of Speex and Dragon Naturally Speaking software. Appendix B includes MATLAB codes used in the simulation.

II. VOICE OVER INTERNET PROTOCOL

Telephone traffic has evolved over the last decade or so. First of all, a change from analog to digital telephony has well been established in most countries worldwide. Second, there is an increasing trend towards the use of internet telephony, also known as Voice over Internet Protocol (VoIP). VoIP is the transmission of conversations using a packet switched network, which is usually based on the transmission control protocol/internet protocol (TCP/IP) suite. There are plenty of reasons for switching from traditional voice transmission to packet telephony networks. First of all, voice transmission over the Internet can be cheaper than that over traditional telephone networks [4]. Second, it provides a handful of new opportunities and applications to its users; these new features are almost impossible to implement without the use of a packet switched network. These, of course, are not without drawbacks, such as the limited availability of the VoIP network, the inability to be used for emergency applications like 911 calls, and the consumption of network resources.

A. VOIP OVERVIEW

There are many different ways in which two or more users can be connected to a VoIP network, but the main concept of interconnection remains pretty much the same. First, a call control protocol is used to initiate the connection between the two users. After the connection has been established, the users can talk. As shown in Figure 1, the voice of one user is digitized, compressed and then packetized before being sent through a wired or wireless communication channel to the other user. At the other end, the opposite procedure is followed: the received packet is depacketized, decompressed, converted to analog form and then played back to the user. In order for the conversation to be natural, the same procedure must be followed in both directions so a full duplex communication is established. The simplest implementation that one can have includes two devices running a VoIP application separated by the internet. In order for the two users to communicate (voice), a logical connection must be initiated by a call control protocol. Then they have to be connected to local area networks, which in turn are

connected with a gateway router to the internet. A simplified sketch of the abovementioned configuration can be seen in Figure 2.

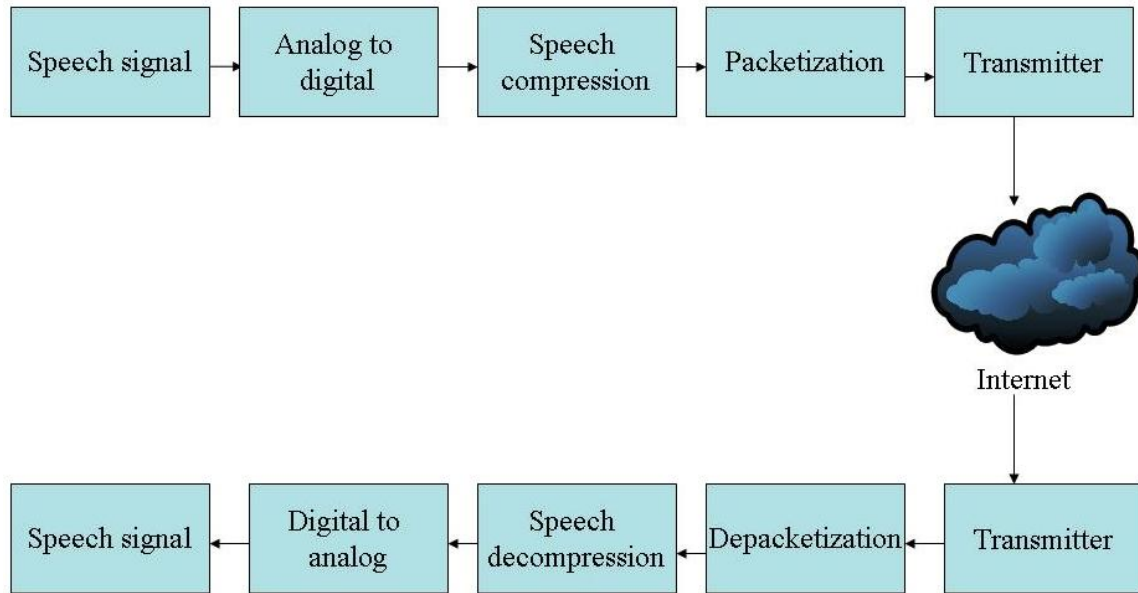


Figure 1. Basic VoIP Communication System

The gateway is an essential component of a VoIP interconnection and implements the following. First, it provides Public Switched Telephone Network (PSTN) and VoIP signaling interfaces, combined with signaling conversion function between the two interfaces, if both PSTN and VoIP networks are in the signal path. Second, it provides a media interface for VoIP and PSTN as well as a media transformation function in the case where both PSTN and VoIP are used. A gateway in general shoulders the responsibility for the connection management from media exchange and signaling flows and thus is an important part of the connection. [11]

With the advances of wireless communications and the increasing use of WiFi (Wireless Fidelity) and satellite networks, there is an increased trend to have a part of one's network implemented through a wireless or satellite link. The implementation can be seen in Figure 3 with the various combinations of wired and wireless media between the two end users being numerous [12]. After the voice is compressed and packetized at the transmitter, it is depacketized and decompressed at the receiver. The voice packets

travel exclusively in IP form from end to end. Inside the internet, the conversation can follow any possible path including wired and wireless connections (e.g., satellite or microwave links) as shown in Figure 3.

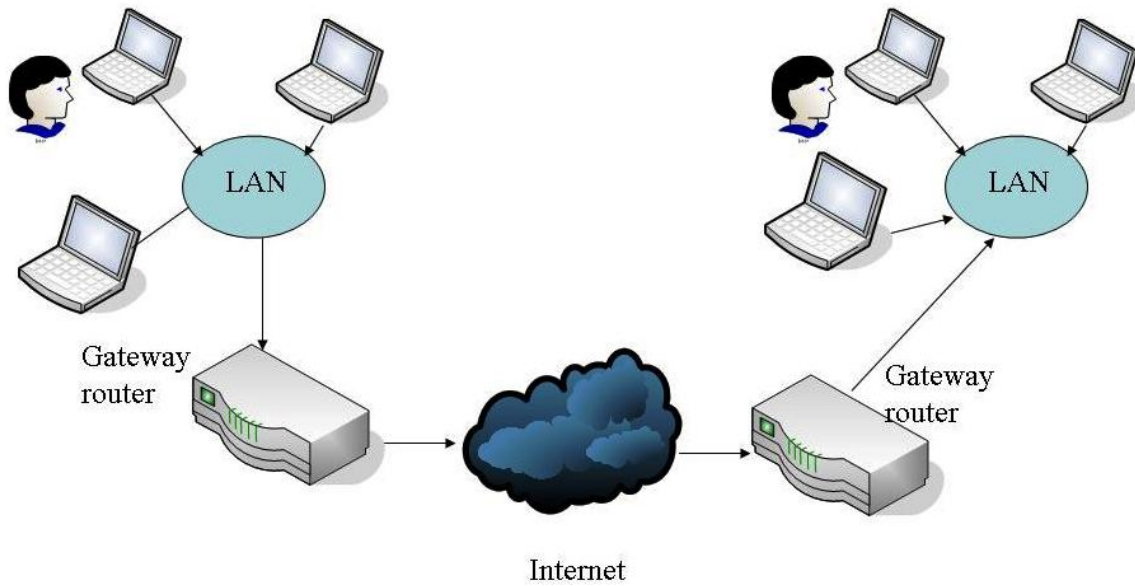


Figure 2. IP to IP VoIP Implementation Interconnecting two Different LANs with Internet

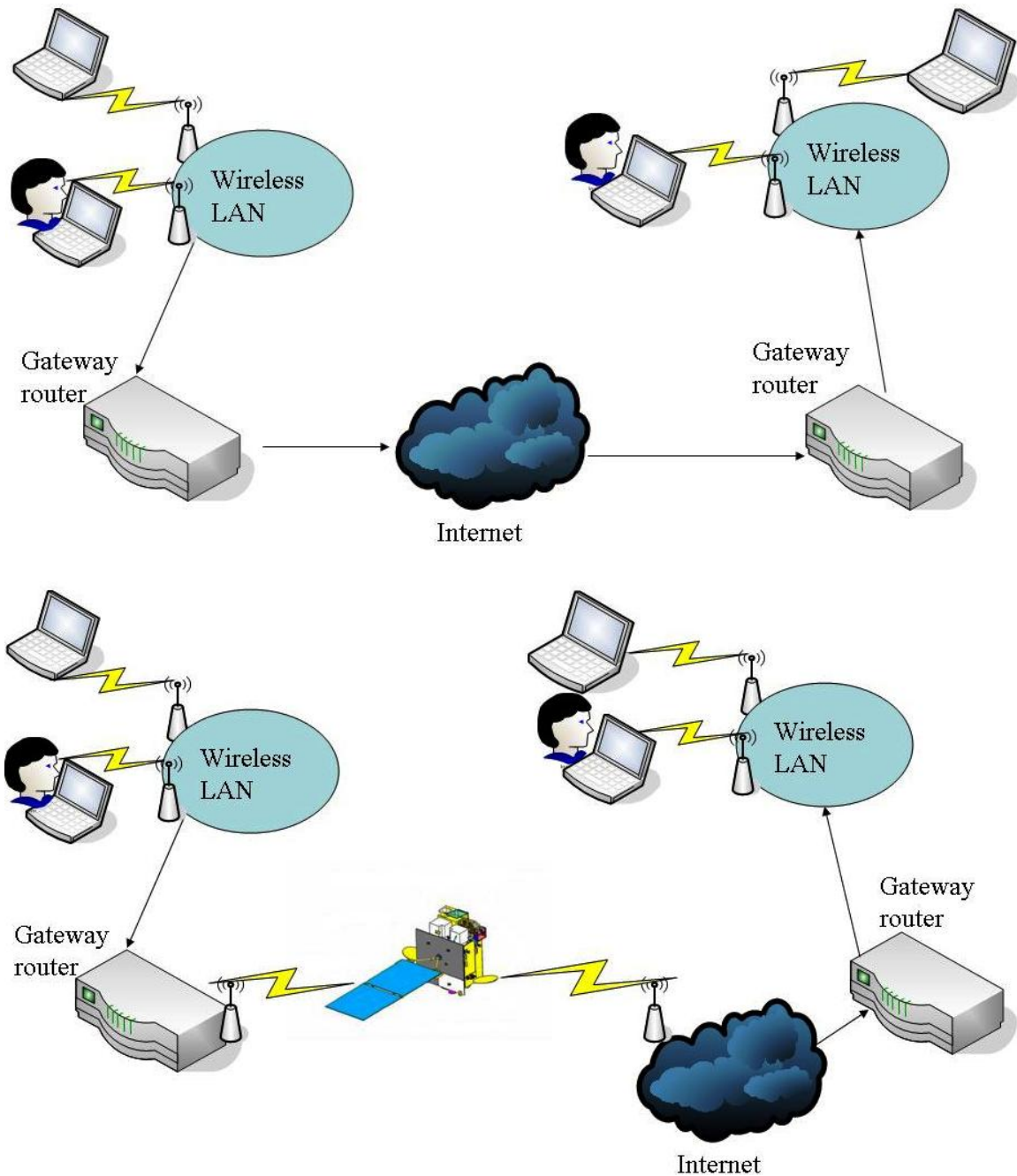


Figure 3. IP to IP with Wireless or Satellite Implementation Included on the Interconnection of two Separate LANs

Voice transmission through a packet switched network is not limited to Internet users only. A VoIP user can communicate with a user of the PSTN given that there is a private branch exchange (PBX) connected to the PSTN [13]. A PBX is a private telephone switch used within large organizations and it can support numerous local loops

as well as provide such functions as call return, teleconferencing, and voice mail [14]. Furthermore, in order for a traditional telephony user to interconnect with a VoIP network, there must be a way to convert signals from analog to digital (IP) form. In order to convert digital IP signals to analog and vice versa, an IP-PBX with capability to accept both traditional telephony and TCP/IP signals can be used. Another way would be the PSTN users to connect to PBX and the PBX to connect to a gateway. [15]

Such an implementation could be very useful when, for example, a VoIP user in Monterey, California wants to call a PSTN user in Greece. Given that the user in Greece has no access to the internet, the user in Monterey can be connected through VoIP to a PBX in Greece and thus establish a local connection instead of an international trunk connection. The interconnection of a VoIP network to a PSTN can be seen in Figure 4.

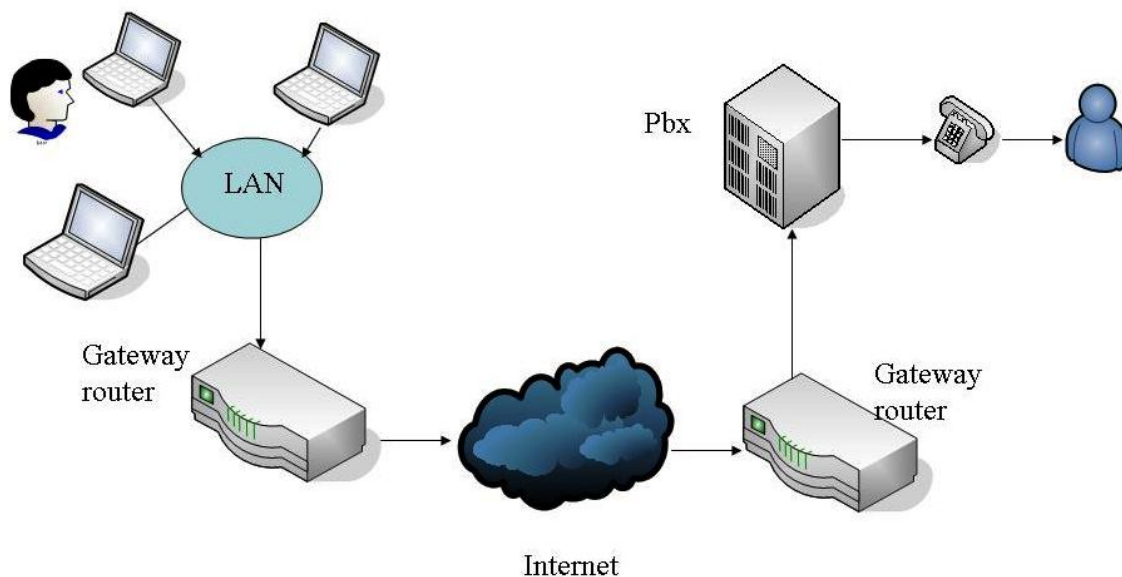


Figure 4. VoIP to PSTN: Interconnection between a VoIP User and a Traditional Telephony User

Finally, the low cost of an Internet call compared to the cost of its legacy competitor has led to the interconnection of two PSTN users with a VoIP network as the interconnecting medium, which can be seen in Figure 5. This implementation could be used by a telecommunications company wishing to connect two remote PSTNs without the cost of wiring (trunk lines) between the two areas.

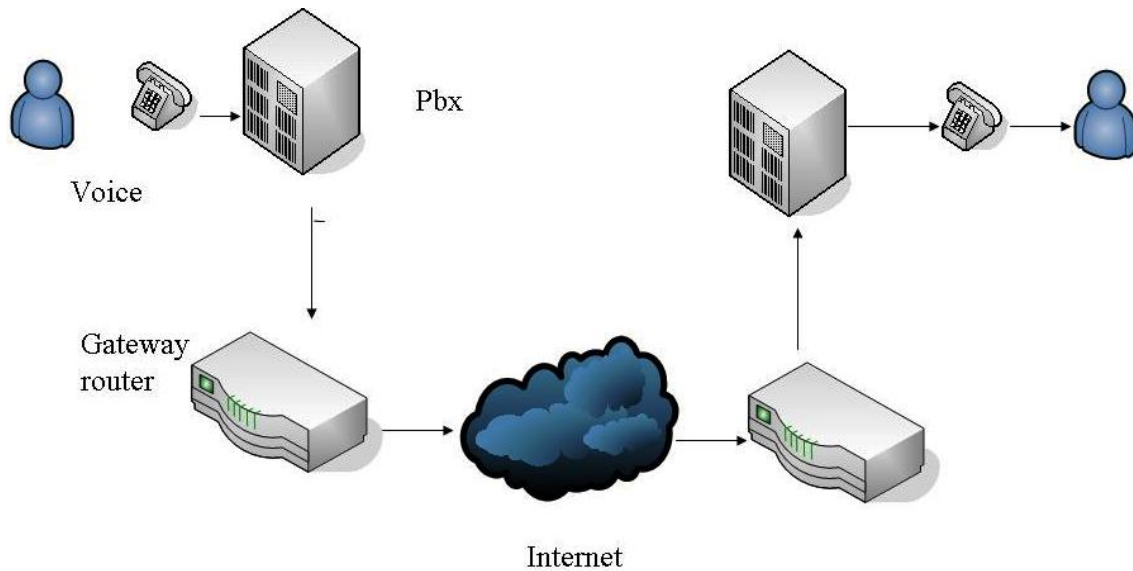


Figure 5. Telco to VoIP to Telco Implementation, Interconnecting two Users of traditional Telephony Using VoIP

1. Protocols

The main protocols used on the Internet today are Transmission Control Protocol (TCP) and Internet Protocol (IP) and associated protocols. Since VoIP uses the Internet as the medium, it also uses the TCP/IP protocol suite. Considering the whole Open Systems Interconnection (OSI) layer system, VoIP is essentially an application running on top of the transport layer. Starting from the bottom to the top of the encapsulation process, any physical and data link layer can be used. IP is the choice of protocol for the network layer. In the transport layer, TCP introduces a large amount of setup delay, which makes it inefficient for voice transmission. The use of User Datagram Protocol (UDP) on the other hand provides for a much faster data exchange but without reliable data delivery. The Real-time Transport Protocol (RTP) and Reliable User Datagram Protocol (RUDP) are the protocols of choice on top of UDP since they are especially created for use in VoIP and media on demand [14]. A layered architecture of a VoIP network is shown in Figure 6 along side the seven-layer OSI model for comparison.

OSI layer	Commonly used protocol in VoIP
Application, presentation and session	RTP, RUDP
Transport	UDP
Network	IP
Data link	Data link
Physical	Physical

Figure 6. OSI Layer and VoIP-used protocols

a. RTP

RTP utilizes the datagram service of UDP and provides two kinds of information. First, it provides a sequence number that helps the receiver reorder the packets. Second, it supplies a timestamp that helps the receiver deal with jitter. The RTP header format together with the network and transport layer headers can be seen in Figure 7. The sequence number field is used in order to provide packet loss detection. The timestamp field provides jitter estimation and synchronization [14]. The fixed part of the RTP header is shown in Figure 7. The additional fields depend on the PT field value. RTP provides the receiver with the tools to reproduce the content but does not provide any control functionality. This functionality is provided by RTCP, which is a companion protocol of RTP and essential for its operation. RTCP provides additional information about the data exchange and the network performance. RTCP packets use a different port number than the RTP stream [16]. RTCP also provides gateway support and source identification in order to allow group conferencing in near real-time [14].

A drawback of the use of RTP is the additional overhead. The IP/UDP/RTP header presented in Figure 7 is 40-bytes long. The typical data payload carried by this packet is two G.729 compressed frames, which is about one half of the header size [14].

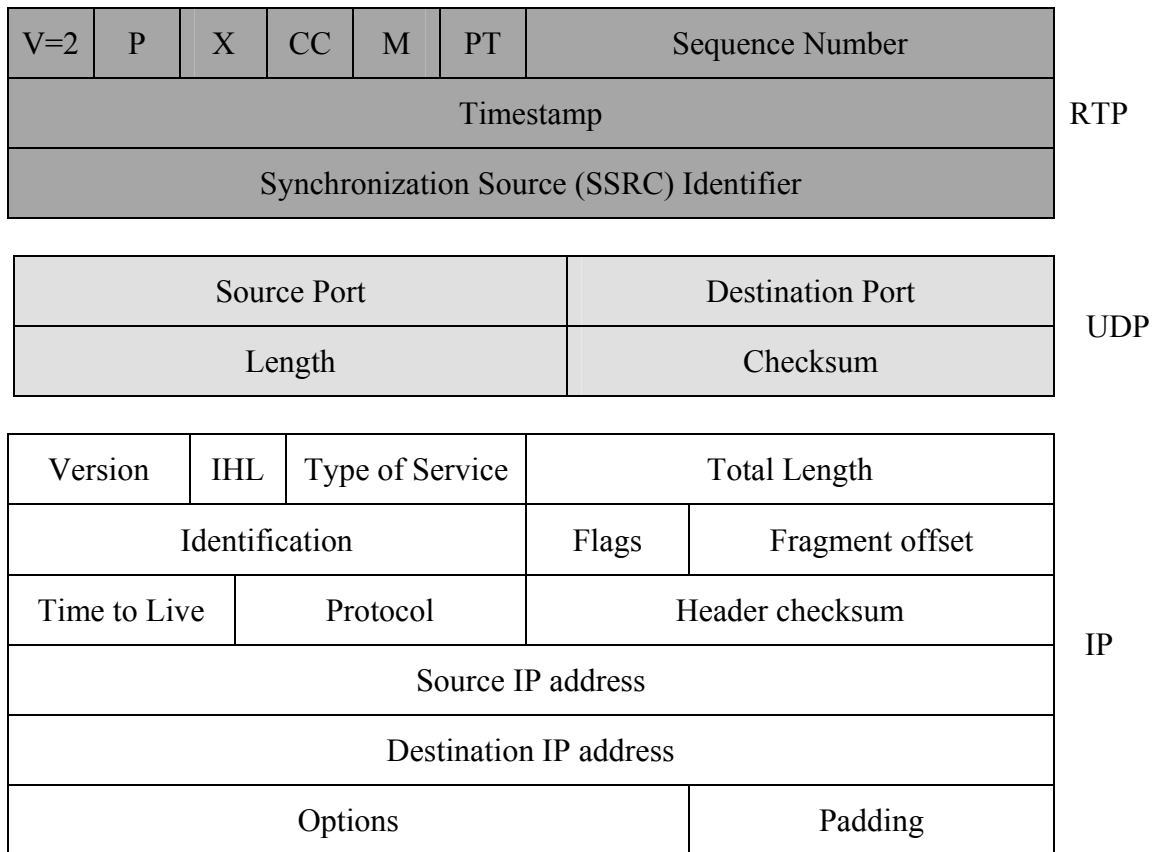


Figure 7. IP/UDP/RTP Header Format

b. RUDP

RUDP is an alternative protocol to RTP. RUDP provides some reliability and survivability to UDP. The reliability is achieved by sending more than one copy of a packet to the user in hopes that one of them will eventually make it to its destination on time. It is able to provide in-order delivery in a reliable way, but, despite a simple implementation, it is also bandwidth consuming. Even though it may require double or triple the bandwidth used, in cases where reliability is a major concern, it is the preferred solution [14]. RUDP rides on top of UDP the same way as RTP, and its header format is shown in Figure 8. The sequence number field is randomly chosen when a connection is opened and is incremented by one for every packet sent. The checksum field uses the same algorithm as TCP and UDP and provides integrity for the header part of RUDP.

SYN	ACK	EAK	RST	NUL	CHK	TCS	0	HEADER LENGTH
SEQUENCE NUMBER								ACK #
CHECKSUM								

Figure 8. RUDP Header Format

2. Signaling in VoIP

Before a communication takes place in conventional telephony, some necessary steps must be carried out. After the caller picks up the phone and dials the called party's number, a signaling protocol is activated in order to find if the called party is available. If the called party is available, it establishes a line of communication for the two parties to talk. The same procedure is followed in VoIP and can be seen in Figure 9. First, the signaling protocol looks for the IP address of the called person and if he or she is available and willing to participate a logical channel is established. The call is established and parameters like voice coding, session protocol and capability exchange are negotiated between the two users. After that, the two users can talk. The signaling protocol is still present and monitors the quality of the call and waits for the signal to terminate the call. There are two signaling protocols widely used in the market. The older of the two is H.323, which is an International Telecommunication Union (ITU-T) standard, and the newer is Session Initiation Protocol (SIP), which is an Internet Engineering Task Force (IETF) standard.

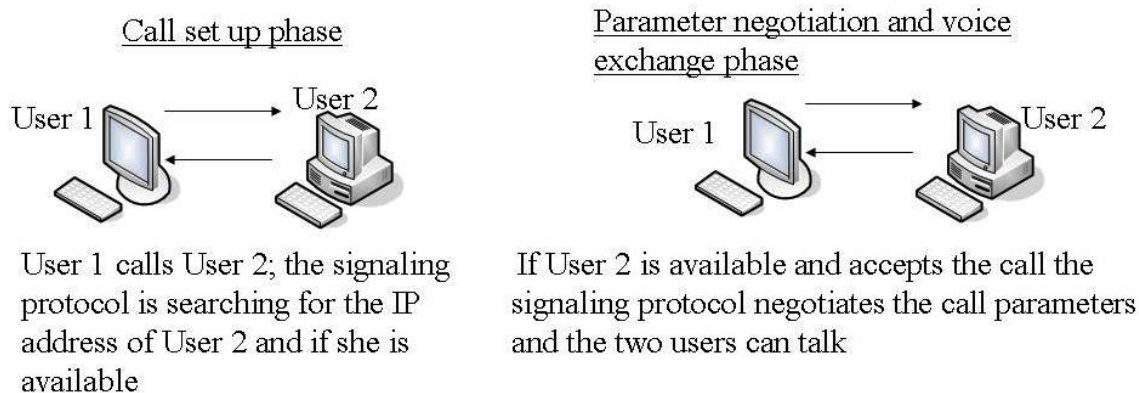


Figure 9. VoIP Signaling Procedure Including Call Setup and Voice Exchange Phase

a. H.323

H.323 was developed in order to allow for transmission of voice and video through the Internet. In addition to signaling, it regulates all aspects of multimedia transport, audio and video codecs, and bandwidth control. H.323 consists of terminals, gatekeepers, and main control units (MCUs). It also consists of protocols and components, such as H.225, H.245 and H.235, which help it integrate the full spectrum of functionalities it can offer. Due to the way it is constructed, it can offer many capabilities and provide interoperability among many vendors. On the other hand, it requires a long call setup delay, considerable overhead and complicated implementation. Despite its substantial disadvantages, it still has a sizeable share of the market [11].

b. SIP

The main idea behind SIP is an application layer protocol capable of doing all the basic signaling functions with simplicity and integrating with all available Internet protocols. In contrast to H.323, it is not an integrated communications system and it needs other protocols to communicate, such as RTP, Real-Time Control Protocol (RTCP), Reservation Protocol (RSVP) and Media Gateway Control (MEGACO). In contrast to H.323, it has to handle only the basic signaling features and in order to provide communication services it must be used with other protocols like RUDP. SIP can cooperate with any transport layer protocol even though it usually uses UDP. The

following signaling aspects are supported: user location, availability and capabilities, and session setup and handling. It is based on a client server architecture in which the clients are very simple elements, which can only send and receive a SIP response. The servers on the other hand are much more intelligent and can be proxy servers, User Agent Servers (UAS), or redirect and registrar servers. The distinction between the different kinds of servers is logical and more than one kind of logical server can exist within the same physical computer [17].

SIP packets can be divided into responses and requests. The packet formats of request and response packets are different. They both consist of a start line, a message header and additional optional data. The request packets are used to locate a user, acknowledge a response, and initiate and terminate a session. SIP responses are used to define the call status, redirect a session or define a client or server error.

The first necessary step to establish a two way communication in SIP is to find the called party. This can be done with the help of the appropriate database server in which the specific user has been registered. The procedure of establishing a call between User 1 and User 2 can be seen in Figure 10.

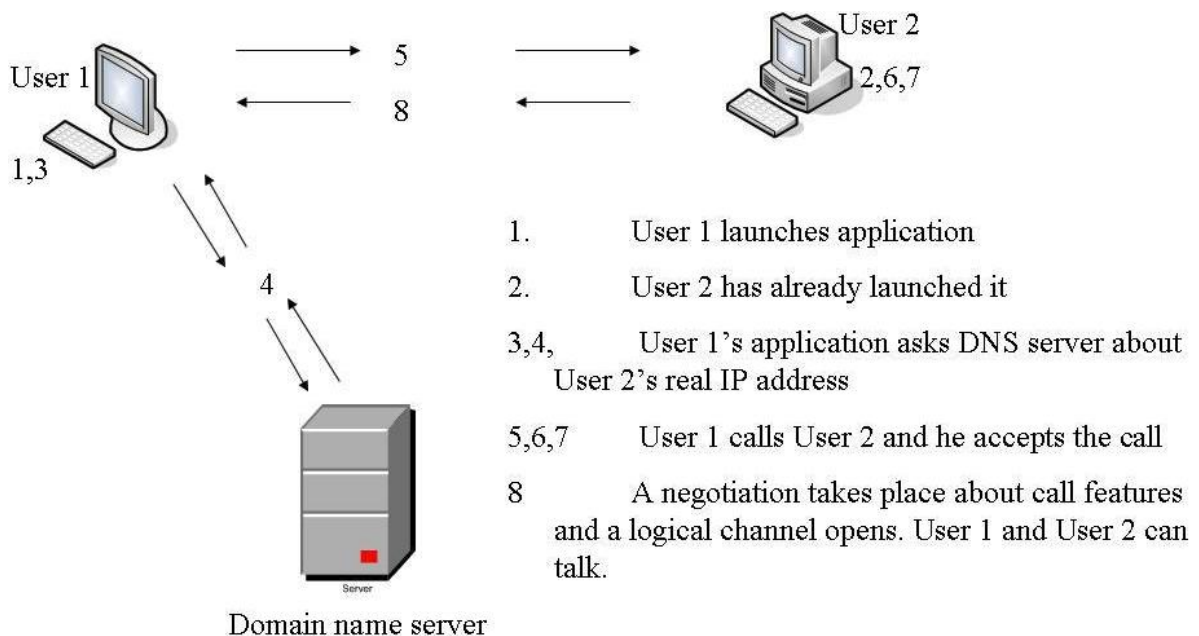


Figure 10. Complete Phone Application Use Including Look Up Phase and Call Setup Phase

In order for the procedure of Figure 10 to be carried out successfully, both users must be registered at the same domain name server. This could be the case if both of them were on the same campus or in the same corporate building. In the case that they are registered to different servers, a follow-up has to be made in order for the SIP messages to be redirected correctly. The main concept of this redirection is the same as the Domain Name System (DNS) redirection on the Internet and can be seen in Figure 11. SIP User Agent 1 is looking for the real IP address of User 2. SIP Proxy Server 1, in which User 1 is logged in, does not know the requested address but redirects the query to Proxy Server 2, which in turn redirects it to a database server. After the two users know the real IP addresses of each other and a logical connection is established, they can communicate directly without the use of the aforementioned servers.

A big advantage of SIP is the wide range of applications that it can implement. A user can not only show where he or she is, but also show if he or she is available or willing to communicate. After a logical connection has been established, the two users can send instant messages, participate in a teleconference, invite more users, exchange files, record the conversation, and many other functions.

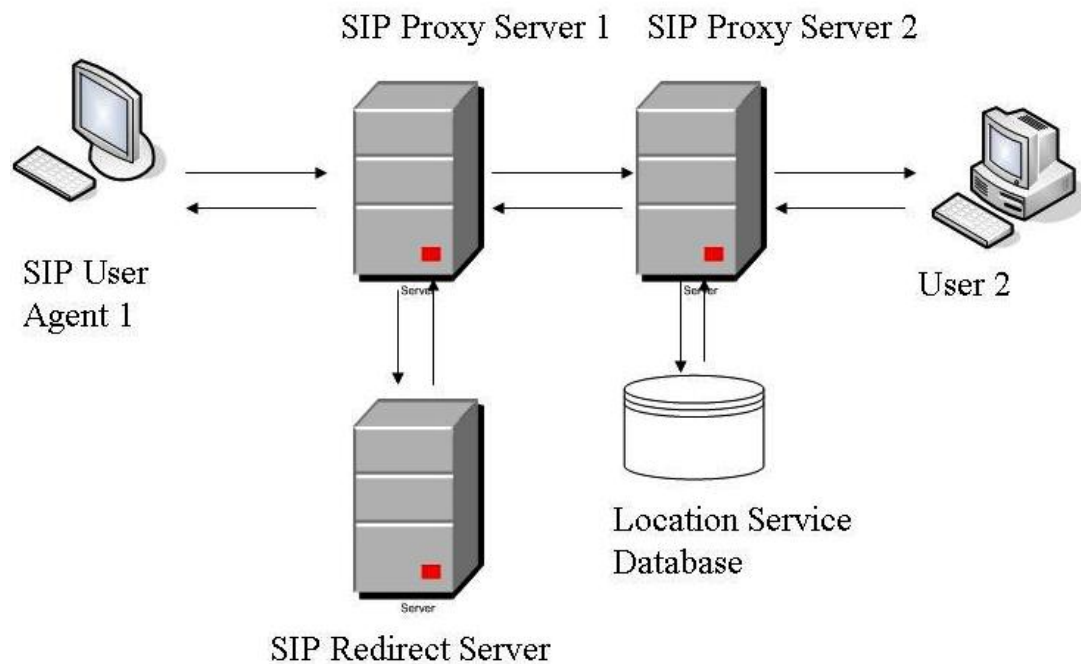


Figure 11. SIP Redirection from User Agent to Proxy Server, Redirect Server and Database Server

B. VOICE CODING

Voice is an analog signal created by air passing through the vocal cords and then the laryngeal, oral, and nasal passages. For voice communication, most frequencies of interest are below 4000 Hz, so a sampling frequency of 8000 Hz is adequate. Each sample is then typically represented by 8 bits, giving a bit rate of 64 kbps, which is considered high for voice transmission. As a result, the voice signal is compressed prior to transmission [18].

1. Speech Compression

The newer compression techniques developed over the last two decades are based on linear-predictive modeling techniques that emulate the human speech production process. Instead of coding the speech waveform itself, the focus is on coding the human vocal system. In this way, instead of sending the waveform or a coding of the waveform, the parameters representing the human vocal system to encode and synthesize speech are sent. More specifically, the excitation source (which is the speech generation) and the vocal tract filter which simulates the modulation of voice as it travels through the vocal tract are sent [19].

Channel vocoders were the first attempt to use this compression scheme, and many different variations of this approach are still a subject of research. Linear predictive coding, Code Excited Linear Prediction (CELP), and some variations of it are presented next.

2. Standards of Speech Compression

a. Linear Predictive Coding

In linear predictive coding of speech, the source is represented by voiced or unvoiced excitations. In general, a linear filter is used to model the vocal tract. The input to this filter is random noise or a periodic pulse, depending on whether the excitation is voiced or unvoiced. Figure 12 shows the model of the human speech production process used by linear predictive coding.

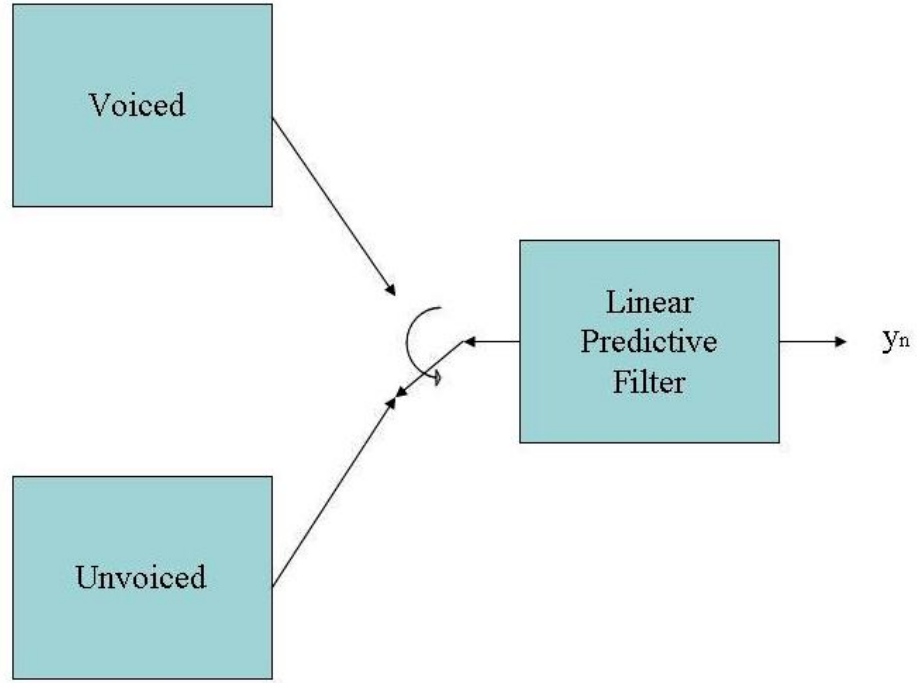


Figure 12. Model for Human Speech Production Process used by Linear Predictive Coding

The output of the filter is given by:

$$y_n = \sum_{i=1}^M a_i y_{n-i} + G e_n \quad (2.1)$$

where e_n is the excitation (also known as prediction error), M is the filter order, G is the filter gain and a_i are the filter coefficients. The linear predictive filter coefficients are obtained by minimizing the prediction error power with respect to the filter coefficients a_i , which is given by

$$E[e_n^2] = E\left[\left(y_n - \sum_{i=1}^M a_i y_{n-i}\right)^2\right] \quad (2.2)$$

where $E[.]$ is the expectation operator [19].

In practice, in order to compress a segment of speech, it is first divided into smaller segments (in the case of Federal Standard 1015 (LPC-10), 180-sample segments are used). This segment is then classified as voiced or unvoiced based on the energy and frequency contained within. After obtaining the pitch and vocal tract filter coefficients, these parameters are transmitted which in turn will be used in the receiver to reproduce the voice. Even though the reproduced speech tends to be unnatural and noise is certainly a problem, LPC is effectively used in applications where compression ratios are of most importance. Linear predictive coding is used in the government standard FS-1015 (LPC-10) which can achieve bit rates as low as 2400 bps [19].

b. Code Excited Linear Prediction (CELP)

Using random noise and periodic pulses as excitation leads to low quality voice reproduction [19]. CELP methods improve the voice quality by using better excitation techniques. The output of the filter is given as

$$y_n = \sum_{i=1}^M a_i y_{n-i} + \beta y_{n-P} + G e_n \quad (2.3)$$

where G is the filter gain and a_i are the filter coefficients. The fundamental harmonic period, also known as pitch, is P and β is a scaling factor. The pitch periodicity contribution is βy_{n-P} and is calculated every subframe [19].

Equation (2.3) can be treated as a cascade of two filters. The first filter extracts the pitch, and the second is a long term formant filter. The excitation is created using the codebook approach so that it is not necessary to extract voicing patterns or pitch. The codebook is generated offline and every time the synthesized outputs are compared with the predetermined codebook to find the best match. CELP is used in Federal Standard 1016 where two codebooks are used as seen in Figure 13. The first codebook, called stochastic codebook, is fixed and predetermined and the second is adaptive. The excitation of each segment is the sum of the adaptive and stochastic codebook outputs. After excitation $e[n]$ is produced, a copy of it is fed back to the adaptive codebook which then adapts to the current segment. In order for the scheme to

provide minimum error between the input and synthesized speech, the codebook indices are scaled using gains. FS-1016 provides very good performance at rates 4.8 kbps and above [19].

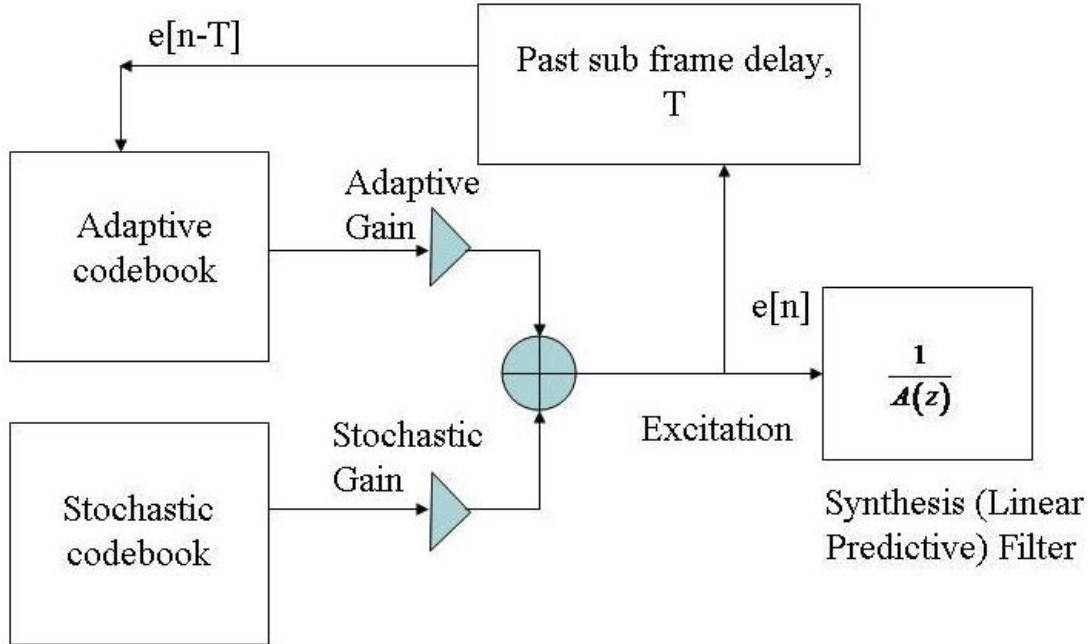


Figure 13. Block Diagram of Federal Standard 1016 Using CELP and Featuring two Codebooks, Stochastic and Adaptive

c. *CELP Variations*

Variations of the CELP approach are used in many commercially used speech codecs. Below are some of them.

1) Low Delay (LD) CELP

LD-CELP is used in the ITU G.728 standard and has a 2.5 ms delay with a 16-kbps bit rate. It uses only a short delay predictor, and the speech segment is 20 samples long and 2.5 ms in duration. The excitation vector is defined using 10 bits [20].

2) Vector Sum Excited Linear Prediction (VSELP)

VSELP was standardized in IS-54 and offers an 8 kbps data rate. It uses 20-ms speech segments and two codebooks for excitation as in FS-1016. It is used in cellular mobile radios in North America [20].

3) Qualcomm CELP (QCELP)

It was standardized in IS-95 and offers data rates in the range of 1-8 kbps. It uses 20-ms speech segments and two codebooks for excitation as in FS-1016. It is used in digital cellular systems in North America [20].

C. VOICE RECOGNITION

Voice recognition is the technology that allows machines to receive, analyze, and act on a speech signal. The action can be converting the speech to text, executing the spoken instructions or responding to the speaker by using synthetic speech.

1. Different Approaches and Constraints

First of all, speech recognition has to deal with all the linguistic constraints that make up human languages. All the grammatical, syntactical, lexical, and semantical rules must be taken into account in order to achieve efficient speech recognition.

The determining factors for speech recognition are the size of the supported vocabulary and the user dependency. Complexity and difficulty of recognition increase logarithmically as the vocabulary size increases. It is also a much simpler task to recognize speech from a specific user for which the system is trained for than being able to recognize speech independently of the user [21].

Speech recognition can be done as isolated word recognition (IWR) or continuous speech recognition (CSR). In the isolated words approach, the system needs discrete speech units as inputs. It requires pauses between words and is a simplified approach that demands cooperative users. IWR is suitable for some applications, yet the most used technique today is CSR. CSR is more complex and uses spontaneous, natural speech. Except for the linguistic constraints, it has to deal with temporal boundaries and

coarticulatory effects [21]. The two main algorithms used today in speech recognition are Dynamic Time Wrapping (DTW) and the Hidden Markov Model. [21]

2. Dynamic Time Warping (DTW)

Dynamic time warping (DTW) is a simple method and is mainly used in IWR. In order to be used in CSR, the connecting speech method is necessary. Dynamic programming is used to find the minimum cost between nodes and has many other uses except speech recognition. DTW uses a database as reference and compares each utterance, which is usually a word with it

Depending on the speaker and the use of a specific word, a word can have different duration than that in the reference database. In order to provide the best match, DTW performs time contraction and expansion before comparing the word with the database sample. A more efficient approach uses energy measures together with the time contraction and expansion of the word. In this way a more efficient weighting of the expansion is done, yielding more accurate results [21].

3. Hidden Markov Model (HMM)

HMM is used since 1970 and is more efficient method than DTW. It solves the speech variability problem more efficiently. HMM uses two stochastic processes, one measured and one hidden in order to simulate the lost observations. From these two observations, only the observed process is used to extract information and characterize the process. It is essentially a state machine with states representing the features of vocalization [21].

In order to achieve speech recognition, the HMM approach requires two stages: training and recognition. During the training phase, a database is created with the statistical features of each word and the way these features are statistically associated. During the recognition phase, each word to be recognized is contrasted with every database element. The features of each element are compared using the HMM algorithm. The best match is the recognized word [21].

An HMM is defined by the number of transitions. A six stage transition can be seen in Figure 14. For every state jump, a sequence observation is generated and another one is discarded [21].

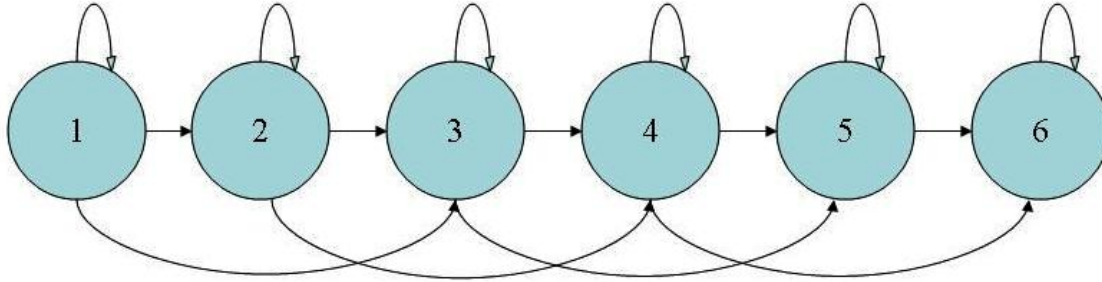


Figure 14. HMM with Six States Labeled as Integers

HMM and DTW are the methods mainly used for speech recognition today. HMM is complex but has the benefits of CSR capability and extended vocabulary use with specific and non-specific users. DTW on the other hand is simpler, limited to IWR with limited vocabulary, and used mainly for specific users [21].

D. QOS

Quality of service in packet switched networks is referred to as the means to provide an assured bandwidth, or stated otherwise, it is a way to define the network's performance. For this thesis, the term Quality of Service (QoS) will be used to describe the performance of the VoIP schemes. Some of the ways to define it are by measuring network availability and voice quality. Definitions, as well as the factors affecting QoS will be discussed in the remainder of the chapter.

1. VoIP Availability

For many years the PSTN companies have been advertising their four 'nines' of availability as their main advantage over VoIP and cellular telephony. What they really mean is that their network (from PBX to PBX) availability is close to 99.99%; if the availability of home appliances has to be accounted for, the total availability would be

much lower. In order for VoIP to fully replace the traditional telephony, it has to get closer to its competitor's level of availability. Availability according to [22] is given by

$$AVAILABILITY = \frac{MTBF}{MTBF + MTTR} \quad (2.4)$$

where MTBF is the mean time between failures and MTTR is the mean time to recover. Even though there are many measurements in regards to Internet availability, very few exist for VoIP availability. According to [23], the following results are available for VoIP. Call success probability is 99.53% and overall network loss is 0.56%. If one considers a total network loss as the failure of 5% or more packets, then the overall network loss falls to 2.52%. Network outages with eight or more packets in a row are 0.56% and may end up with up to forty seconds of consecutive speech loss. The probability to abort a call due to network outage is 1.53% (i.e., a user hangs up the phone after hearing nothing on the other side). Finally, if one considers a call as successful if it gets through and it is not aborted, then the total probability is 98%, which is far from the promised four 'nines' of the PSTN providers.

2. VoIP and E.911

Except for the aforementioned limited VoIP availability, there are some other limitations of VoIP when it comes to emergency calls. Emergency calls through Internet telephony are regulated by the Federal Communications Commission (FCC) [24]. The limitations of VoIP in emergency calls are the inability to support caller identification (ID), caller location, and call back information, since a VoIP user can call from any place he wants, provided that he or she has an Internet connection. The next limitation has to do with the E.911 service. E.911 services are provided to users of mobile telephony and VoIP, and the physical location of the user is transmitted when they dial the 911 service. The problem arises when VoIP calls, instead of going to the appropriate Public Safety Answering Point (PSAP) authority, are connected to administrative personnel or a call

may not be possible due to network overload or failure. The measures provided by the FCC are mandatory implementation of the E.911 calling feature as well as the ability to provide caller Identification (ID) and call back information. Additionally, all users must be informed of the limitations of VoIP in regards to emergency calls.

3. Factors Affecting QoS

Voice quality is influenced by a variety of factors, namely packet delay and packet delay variation (also known as jitter), packet loss and, finally, the type and amount of voice compression used. The following is a brief description of all the factors affecting the QoS [14], [25].

a. Packet Delay

Packet delay is of two types: handling and propagation. Handling (packetization) delay is the amount of time it takes for a speech signal to be processed by the computer's hardware before it is transmitted to the medium. Propagation delay is the amount of time it takes a signal to travel from transmitter to receiver and is dependent upon the medium used. The effect of the total delay is annoying to users in a conversation. There are limits on the total delay depending on the type of communication [24]. A delay of 100 ms or less cannot be perceived by the human ear, a delay of 150 ms is perceivable but the level of the conversation is acceptable, and beyond that the speech quality is not acceptable except in specific circumstances like in a satellite transmission where a delay of 400 ms is still acceptable since one cannot avoid it due to the large physical distances involved [25].

b. Packet Delay Variation

Packet delay variation or jitter is the variation in delay between consecutive packet interarrival times. It is an effect of packet switched networks and it can be more annoying than the packet delay itself since its effects vary over time. In order to compensate for jitter, a VoIP network has to establish a jitter buffer that reorders the packets and waits until enough packets have arrived to be played back. The longer the

jitter buffer, the longer the delay and the less the jitter perceived by the user. Jitter buffers can be a fixed length or can vary in order to handle any excessive jitter [25].

c. Packet Loss

Another problem of most packet switched networks is packet loss. It can happen at any point in the network especially in media that are prone to errors like a wireless medium. When using a connectionless protocol like UDP, the sender cannot be sure which packets have been received by the other side. Also, when a packet is delayed, it is sometimes better to drop it rather than increasing the buffer size so much in order to account for the jitter. This is because by increasing the buffer size, we also increase the delay in the conversation. One of the techniques used in order to account for the loss of a packet is to replay the packet received before the one that was lost. In that way, instead of short periods of silence, one listens to the voice that seems slightly distorted. A packet loss of 3% is generally considered as the maximum tolerable amount of loss [14].

d. Type of Codec

The type of codec (voice compression schemes) used is a critical part of the VoIP network. It determines the bit rate needed, the complexity of the encoder, segment length (and thus the handling delay induced by the coder), and finally the quality of the received speech. Some of the most popular codecs used today are listed in Table 2. In a VoIP network, codecs may be used in tandem if, for example, after a VoIP network, there is a PSTN, which uses a different codec. In this case, there is degradation in the received speech quality, which depends on the codec and the number of tandem encodings [13].

4. Clarity of Received Speech and Methods of Measuring It

The final and most effective measure on any telephone call is how the user perceives what he or she hears which means how clear and undistorted the sound from the other user is. It is obvious that such a measurement is subjective. It is subjective to the user's background, mood, and attitude. The metrics for the expected speech quality have been set by the PSTNs over decades of use, but with the introduction of packet telephony

some new factors affecting the quality of speech have to be taken into account, such as packet loss, jitter, and silence compression. In traditional PSTNs, the signal-to-noise ratio, the intermodulation and harmonic distortion, and the bit error rate (BER) are measured in order to determine the quality of the communication. These metrics are insufficient for use in packet telephony since, even though in some cases, excellent metrics can result in bad speech quality. This is due to the factors affecting VoIP quality, such as packet loss, jitter and echo, which can not be completely described from the above mentioned methods [26].

CODEC	Frame size (msec)	Bit rate (kbps)
G.711	0.125	64
G.723	30	5.3-6.3
G.726	0.125	32
G.728	0.625	16
G.729	10	8-11.8

Table 1. Popular Codecs with Their Frame Size and Bit Rate

There are two general categories of measuring the quality of speech: subjective and objective. In the subjective category, the effort is concentrated on investigating how people perceive a given speech sample. On the other hand, in objective measurements, mathematical formulas are used in an effort to get results as close as possible to subjective tests. Mean opinion score (MOS) is one of the standard methods used to subjectively measure speech quality. It uses a large volume of human opinion scores on a specific speech sample to measure its quality. The users rate the speech from 5 which is (excellent) down to 0 (bad). Alternatively, the users can grade the speech depending on the effort required to fully comprehend the meaning of the speech sample. MOS has been standardized by ITU-T as a telephone speech quality metric [27], [28].

In addition to being subjective, MOS is also very expensive and it cannot be used continuously to measure the effectiveness of a network. The latter drawback led to the implementation of automated, objective measurements. The first implementation was the Perceptual Speech Quality Measurement (PSQM), which basically considers both human perception and the subjective nature of quality. The difference between the original and the distorted voice is measured. The input sound is compared with the output on the frequency domain based on how humans perceive speech. The metric used is from zero to infinity with zero representing total match and infinity representing no match at all. Despite its efficiency, it is just a way to simulate MOS results and it still may not account for delay, jitter, multiple talkers and low bit rate coders. Some subsequent alternative techniques include Measuring Normalizing Blocks (MNB), PSQM+, and Perceptual Evaluation of Speech Quality (PESQ), with the latter having the best results among its competitors. A more detailed presentation of both subjective and objective speech quality measurements can be found in [26].

E. SUMMARY

VoIP is an emerging technology that is striving to compete with if not to supersede the traditional PSTN telephony. An overview of the main concept and the major technical aspects has been presented and voice coding has been introduced. QoS from the VoIP point of view has been discussed along with the main factors affecting it. Next, the focus will be on the medium in which the voice travels and especially on the wireless channel.

III. WIRELESS NETWORKS

Wireless networking is an increasing trend in digital communications today. There is no physical cabling required, and the installation cost is significantly lower than the cost of a wired installation. The topic of discussion for this chapter is the use of wireless channels with emphasis on digital transmission of voice.

A. DIGITAL WIRELESS COMMUNICATION NETWORK

In computer networks, information is transmitted in digital form, i.e., using a long sequence of ones and zeros. The transmission can either be done through guided or unguided (wireless) media. An example of guided media is a copper cable, and an example of unguided media is the transmission of electromagnetic energy through the atmosphere. In both cases, the main system functions before the transmission of the signal and after its reception remain the same and can be seen in Figure 15. Before transmission, the signal must be encoded (compressed), then channel encoded, pulse shaped and modulated, and then transmitted through the medium. In the receiver, the reverse procedure is followed [29].

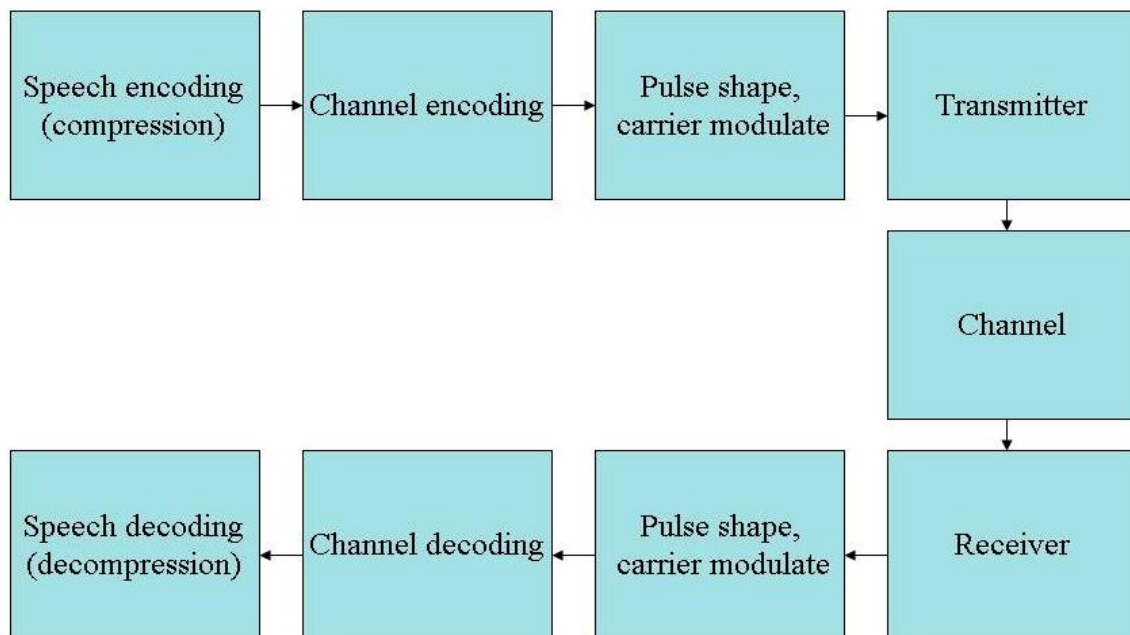


Figure 15. Block Diagram of Digital Voice Communication System

Combining a receiver and a transmitter on the same circuit produces a transceiver which is the most common implementation. In order to create a wireless network, one needs to interconnect two or more wireless nodes. The interconnection of the nodes is achieved over a predefined radio frequency band. A typical wireless network can be seen in Figure 16 where four wireless nodes are interconnected to form a wireless network.

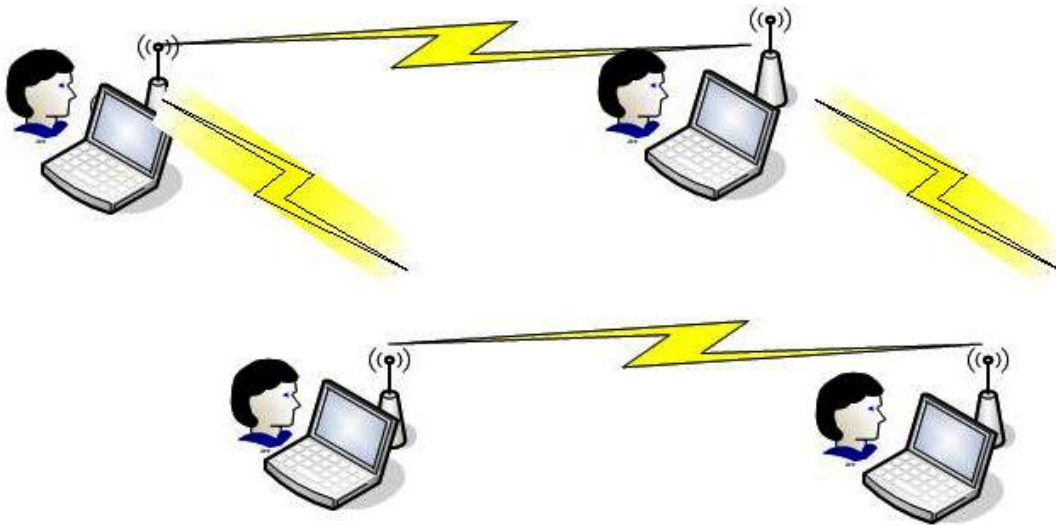


Figure 16. Implementation of a Wireless Network Using a Sum of Interconnected Nodes

Interconnection of the nodes using radio frequency means that all the participating nodes have to use the same spectral band. This is a very important aspect of the wireless medium since there must be a means to manage access to the medium. Without management of medium access, all nodes would be using the medium simultaneously, causing collisions that degrade quality. This is a task commissioned to the medium access control (MAC) part of Layer 2 in the OSI model.

B. CHANNEL

In order to analyze and describe the effects of the medium between transmitter and receiver on the system performance, a channel model is used. Essentially, it describes the effects of the physical path on the communication performance. Two kinds of channels are used: wireless and wired. The subject of this section is the wireless channel.

Wireless channels are affected by a number of factors including the physical distance, which causes path loss. They are affected by Doppler and multipath delay spread, interference, and noise level. The propagation parameters depend among others on the atmosphere, terrain and antenna characteristics. These factors are random variables and as such the channel behavior can only be characterized in statistical terms.

1. Attenuation and Noise

In order to fully recover the signal at the receiver's end, two conditions must be met. The signal strength must be sufficient to be detected by the receiver's circuitry and must be higher than noise. The main signal degradation in wireless transmission is due to attenuation. Attenuation is the degradation of the signal over distance due to the spread of energy over a continuously larger surface area. The free space loss is expressed as the ratio of transmitted power, P_t , to the received power, P_r :

$$\frac{P_t}{P_r} = \frac{(4\pi d)^2}{\lambda^2} \quad (3.1)$$

where d is the distance between the transmitter and receiver antennas and λ is the wavelength of the carrier signal [30].

Even though the free space loss model is adequate to describe a satellite link, it is inadequate for any other form of digital communication network. In cases like a mobile radio channel, where there is more than one path between the transmitter and the receiver, it is more accurate to use the two ray model, which uses optical geometry laws. The two ray model assumes that only two rays arrive at the receiver, the direct signal and a signal reflected from the ground. According to the two ray model, the path loss is given by

$$\frac{P_t}{P_r} = \frac{d^4}{G_t G_r h_t^2 h_r^2} \quad (3.2)$$

where G_t and G_r are the gains of the transmitting and receiving antennas and h_t and h_r are the transmitter's and receiver's heights, respectively [31].

In addition to the effect of attenuation, the receiver must detect the signal in noise. Noise can be thermal, intermodulation, crosstalk and impulse. For digital communications, it is easier to represent noise using the ratio of bit energy to noise power spectral density, known as E_b / N_o , which is related to signal to noise ratio (SNR) and bit rate R as follows:

$$\frac{E_b}{N_o} = \frac{S}{N_o R} \quad (3.3)$$

where S is the signal strength, N_o is the noise power spectral density level and R the data rate [32].

2. Multipath

In the case of mobile wireless networks, a loss model that accounts for multiple copies of the same signal due to multipath effects must be considered. The main sources of multipath are diffraction, reflection and scattering. These multiple copies have varying delays and, in some cases, they might be the only signals received, e.g., in the case of nonline of sight (NLOS) reception. An illustration of a scenario causing multipath can be seen in Figure 17 where three different copies of the same signal are received by the receiver: a direct signal, one reflected, and one diffracted. [33].

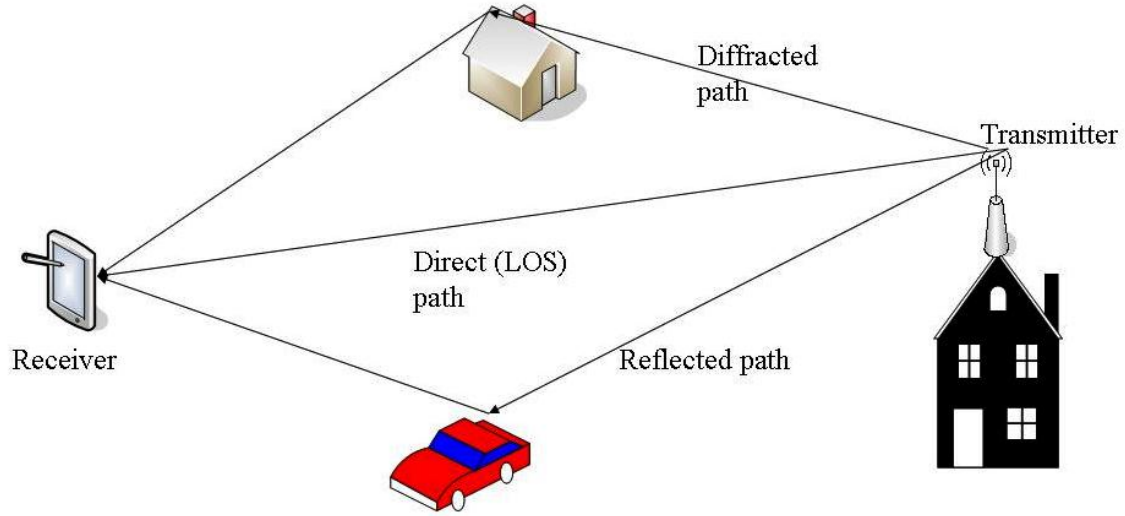


Figure 17. Schematic of Multipath Channel Where the Signal Follows a Direct, a Reflected and a Diffracted Path

When multipath is observed over small distances, rapid changes to signal strength are common. The effect is known as small scale fading and is used to characterize signal variations on a small scale in both distance and time. These changes are due to multipath and Doppler Effect. Doppler Effect is the change in frequency on the received signal due to the relative movement of transmitter and receiver. The result of multipath and Doppler is multiple copies of the same signal with different phase values, overlapping with the signal from the adjacent bit period, which leads to intersymbol interference (ISI). The signal strength can vary even though the transmitter to the receiver distance remains the same [31].

3. Fading

There are two basic models used to simulate fading channels, namely Rayleigh and Rician. In a Rayleigh channel, no path is considered dominant. Signals from different paths having different phase and similar signal strengths are received to produce a Rayleigh distributed signal. The probability density function (PDF) of a Rayleigh random variable is given as

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0 \quad (3.4)$$

where σ^2 is the average received power and r is the signal magnitude. Rayleigh fading reflects the worst case scenario usually found in heavily built urban settings [31].

In the Rician distribution, on the other hand, a dominant line of sight (LOS) signal is present so that the resulting amplitude can be modeled with a Rician PDF as follows:

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + K^2}{2\sigma^2}\right) I_0\left(\frac{Kr}{\sigma^2}\right), r \geq 0, K \geq 0 \quad (3.5)$$

where I_0 is a zero order modified Bessel function and K is the ratio of dominant path power over remaining path power. Dominant power path is usually the LOS path. For $K = \infty$ the channel is additive white Gaussian noise and for $K = 0$ it is a Rayleigh channel. Rician fading is a usual case in indoor or open space outdoor scenarios [31].

A measure for frequency and time dispersion is given by coherence time and coherence bandwidth. Coherence time T_c is the time required to decorrelate two time domain samples. If the time separation of the two signals is smaller than T_c , the signals will be affected similarly by the channel. Coherence time is given by

$$T_c \propto \frac{1}{f_d} \quad (3.6)$$

where f_d is the Doppler shift. When the transmitter and receiver are moving relative to each other, the frequency of the receiver carrier f_c is Doppler shifted by a frequency f_d as follows:

$$f_d = \pm \frac{f_c}{c} v \cos \psi \quad (3.4)$$

where ψ is the angle between the direction of the radiation and relative motion, c is the speed of light, and v the relative velocity of the transmitter and receiver [31].

Coherence bandwidth B_c on the other hand is the frequency separation required to decorrelate two frequency domain samples. If the frequency separation of the two signals is smaller than B_c , then the signals will be affected similarly by the channel. Coherence bandwidth is given by

$$B_c \propto \frac{1}{\sigma_\tau} \quad (3.7)$$

where σ_τ is the rms delay spread of the channel [31].

Depending on the amount of time delay spread, channels can be classified as flat or frequency selective, and depending on the amount of Doppler spread, they can be classified as fast or slow fading [33].

C. MODULATION AND CHANNEL CODING

Two important factors that determine a successful transmission of digital data are modulation and channel coding. With modulation an attempt is made to use the wireless medium as efficiently as possible and with channel coding an effort is made to eliminate the transmission errors.

1. Modulation

In order to transmit a signal through a channel, the signal frequency needs to be shifted to a spectral band appropriate for transmission. This shift is achieved using modulation, which is the alteration of the carrier's characteristics according to a modulating wave. The final result of modulation, apart from frequency shifting, is the addition of information to the carrier signal [34]. In wireless networks, the most widely used modulation schemes are Binary Phase Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), and Quadrature Amplitude Modulation (QAM) [33].

In BPSK, the values of 0s and 1s are represented by two alternating phases of the signal. In QPSK, four different phases of the same carrier signal are used to represent two bits for every transmitted symbol. Finally, in QAM, both phase and amplitude are changed to give the ability for even more bits per transmitted symbol.

The constellation diagrams of BPSK, QPSK and 16QAM can be seen in Figure 18. The BPSK constellation diagram has carrier phases 0 and $\pi/2$ and the QPSK has $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$. The distance between adjacent points is the same and equal to $\sqrt{2E_s}$. The ability to correctly retrieve a bit without error in the receiver is dependent on the distance between the constellation points. By comparing BPSK and QPSK, we notice that it is easier for a receiver to detect a BPSK signal correctly than it is to detect a QPSK signal. Since the distance between points is the same for BPSK and QPSK, all points have the same probability of detection. On the other hand, the occupied bandwidth increases as the represented points per dimension on a constellation diagram increases. This means that BPSK is less bandwidth efficient than QPSK, which in turn is less bandwidth efficient than a 16QAM.

The 16QAM constellation diagram reveals the fact that not only the phase but also the amplitude varies depending on the transmitted symbol. The energy as well as distance between the points is not the same. This is the reason why different symbols have different probabilities of detection. Despite the fact of different probabilities assigned to each symbol, M-ary QAM is the most bandwidth efficient constellation of the three aforementioned modulation schemes.

The probability of bit error rate for BPSK and QPSK is given as

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \quad (3.8)$$

where E_b/N_o is bit energy over noise density power per Hz. Bit error rate can be improved by increasing the energy. The average bit error rate probability for QAM is given as

$$P_e \cong 4\left(1 - \frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{2E_{\min}}{N_o}}\right) \quad (3.9)$$

where E_{\min} is the signal energy of the lowest amplitude and M is the order of modulation. As the order of the modulation becomes higher, the modulation becomes more bandwidth efficient and more susceptible to errors.

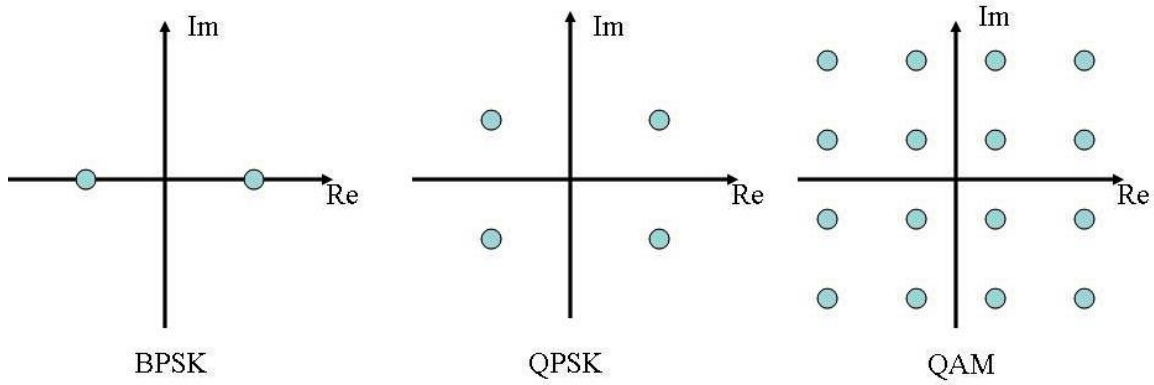


Figure 18. Constellation Diagram of BPSK, QPSK and 16QAM Modulation

2. Channel Coding

In wireless channels, convolutional coding is widely used. The main improvement of convolutional codes over block codes is the introduction of memory to the system and reduced overhead [35] [36].

The parameters describing a convolutional coder are n , k and K . The number of input bits is k and the amount of output bits is n . For example, a rate $\frac{1}{2}$ coder produces two output bits for every one input bit. The term K is the constraint factor that characterizes the memory of the system; an output bit is a function of $K-1$ input bits. The factors n and k can be very small, which makes the code appropriate for use in continuous data streams [33].

An example of a convolutional coder implemented using a shift register can be seen in Figure 19. It implements an $(n, k, K) = (2, 1, 3)$ convolutional coder where the u_n input bit is converted to v_{n1} and v_{n2} output bits. The first output bit is produced from the upper modulo-2 adder and the second from the lower adder.

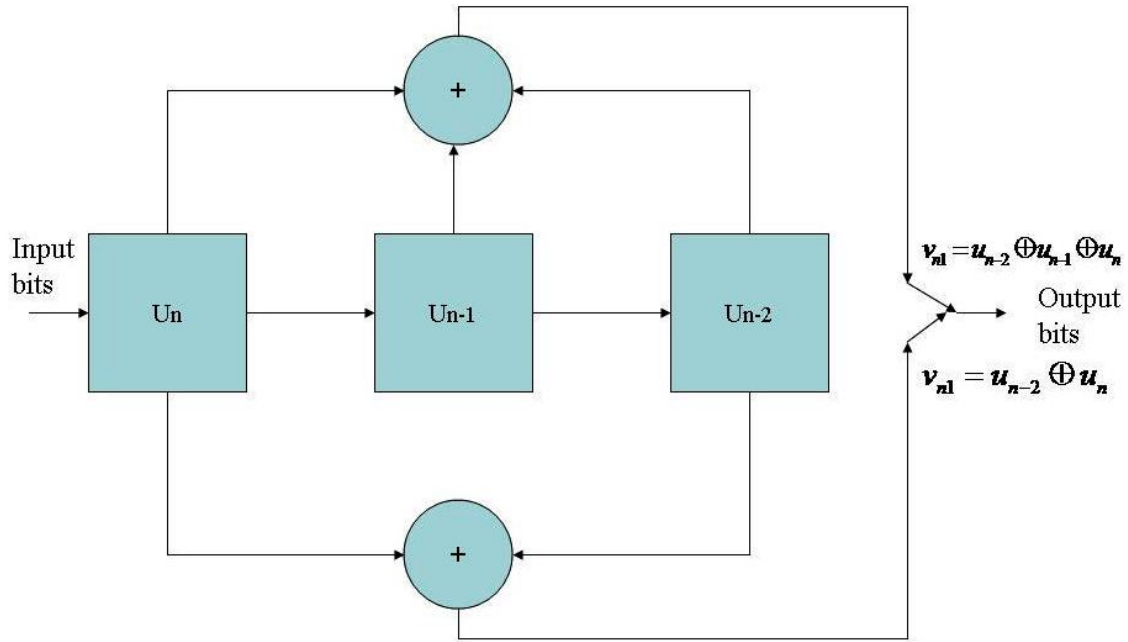


Figure 19. A $(n,k,K) = (2,1,3)$ Convolutional Encoder Implementation Using a Shift Register

D. WIRELESS STANDARDS OF INTEREST

The wireless protocols that are referred to in this thesis are IEEE 802.11 and IEEE 802.16, and are widely used in wireless LANs and backhaul links, respectively. They govern the use of the physical and medium access control so that the higher OSI layers do not have to deal with the details of the medium used and its access. The issues that the physical layer deals with are signal encoding, synchronization, bit transmission, and medium specification. The MAC layer assembles the frame and inserts error detection fields if necessary. It also governs access control to the medium, which is shared between many users.

1. IEEE 802.11

The IEEE 802.11 standard became an IEEE/ANSI standard in 1997 [37]. Several standards followed this main edition with different configurations on the same wireless LAN concept. The distribution system in 802.11 is similar to a cellular system where the

minimum entity is called a basic service set. Each basic service set is governed by an access point, which acts as a relay and as a control station. The access point may be further connected to a distribution system or can be totally isolated like in the case of a small home LAN. An extended service set is a set of basic service sets with access points and interconnection to a distribution system [37].

The MAC part of the 802.11 is common to all physical layer specifications and describes three main areas: medium access control, credible data delivery, and security.

The medium access control part of the standard specifies both a centralized and a distributed way of deciding whether to transmit or not. It uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) with binary exponential backoff. The distributed operation leaves the decision to each station to sense the medium before transmitting. This mode of operation is used to form ad hoc or bursty networks where an access point is not available or its use is impractical. The centralized method on the other hand is governed by a central station (access point) when the data have to be prioritized or have to be delivered as soon as possible. The standard provides the capability to operate contention-free (centralized control) through the Point Coordination Function (PCF) on top of the distributed function provided by Distributed Coordination Function (DCF) [37].

In order to maintain reliable data delivery in an unreliable medium like wireless, there must be a means to provide some reliability without the dependence on higher layers. Reliability is provided by sending an acknowledgment to the source station (two way handshake). For increased reliability, a four-way frame exchange in which a station asks permission to transmit and if permission is granted, it transmits and then waits for acknowledgment is used [38].

Security in 802.11 was initially provided through Wired Equivalent Privacy (WEP) which was essentially a way to protect data from passive eavesdropping. Due to the many weaknesses and vulnerabilities of the system, the 802.11i task group established the Wireless Fidelity (Wi-Fi) protected access protocol, which is an improvement over WEP [38].

The IEEE 802.11 physical layer has many different implementations. The initial 802.11 defined three physical layer units, one infrared at 850nm and 950nm and two radio units operating in the 2.4GHz Industry, Scientific and Medical (ISM) band. The data rates were defined at 1 and 2 Mbps. The 802.11b standard uses Direct Sequence Spread Spectrum (DSSS) at 2.4 GHz with data rates of 1, 2, 5.5, and 11 Mbps. IEEE 802.11a was the next standard with the frequency of operation in the 5.2-GHz and 5.8-GHz bands using Orthogonal Frequency Division Multiplexing (OFDM) and BPSK or QPSK modulation, achieving data rates of 6, 9, 12, 24, 36 and 54 Mbps. Finally, the 802.11g standard with DSSS and OFDM in the 2.4-GHz band provides data rates of 1, 2, 5.5, 6, 11, 12, 18, 24, 36, 48, and 54 Mbps. The main characteristics of the 802.11 physical layers are summarized in Table 2 [39].

The distances covered by the IEEE 802.11 standards depend mainly on the modulation and frequency used. They range from 20 meters for the highest data rate of 802.11a and go up to 100 meters for the lowest data rates of IEEE 802.11b and 802.11g, with the bit rate decreasing as the distance increases [33].

The operation of 802.11a (or 802.11g OFDM option) can be seen in Figure 20. First, the signal is compressed and then channel encoded to provide error correction. Data symbols are formed and OFDM is applied to the signal. The benefit of OFDM is increased resistance of the signal to multipath effects. OFDM is implemented by processing the data symbols through an Inverse Fourier Transform (IFFT) in the transmitter [40]. At the output of the IFFT block, a cyclic prefix is added to the output sequence. The pulse is shaped and the signal is modulated before it reaches the transmitter. In the receiver, the reversed procedure is followed [40]. The modulation schemes supported in the IEEE 802.11a standard are BPSK, QPSK, 16-QAM and 64-QAM with convolutional coding to achieve multiple data rates. Interleaving is also used [41].

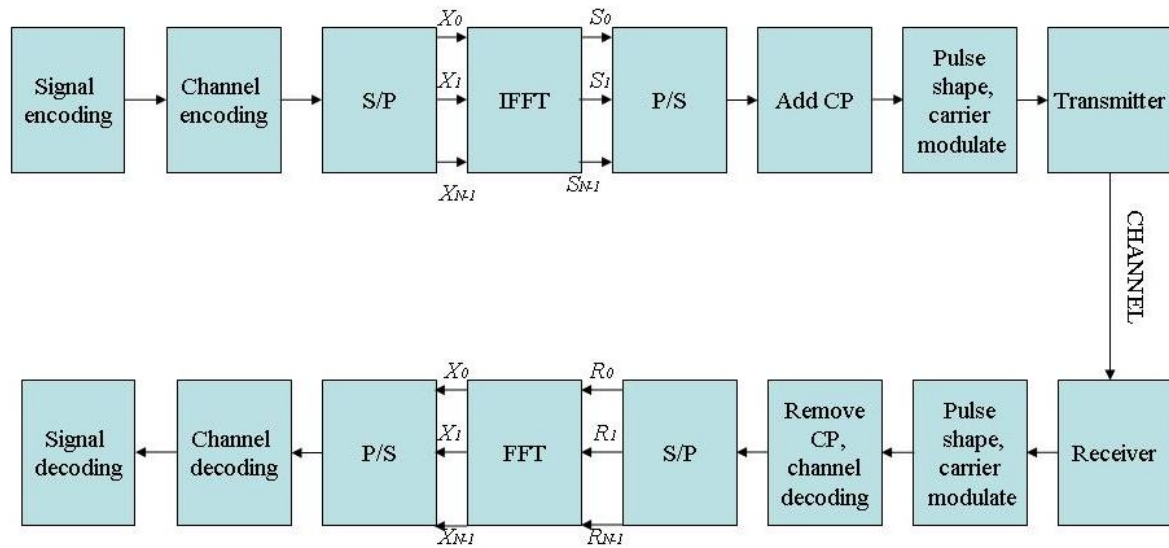


Figure 20. Schematic Diagram of IEEE 802.11a Basic Function Blocks

	802.11	802.11a	802.11b	802.11g
Frequency (GHz)	2.4-2.4835, 850nm(IR), 950nm(IR)	5.15-5.35, 5.725-5.825	2.4-2.4835	2.4-2.4835
Data rate (Mbps)	1,2	6,9,12,18,24,36,48,54	1,2,5.5,11	1,2,5.5,6,9,11,12,18,24,36,48,54

Table 2. Data Rate and Frequencies Used in 802.11 Physical Layer

2. 802.16

IEEE 802.16 is designed to achieve large data rates over distances that can cover a metropolitan area to create a WAN [41]. Subsequent amendments defined network characteristics that can enable a totally mobile network. The standard supports an OFDM scheme (256 subcarriers) as well as an Orthogonal Frequency Division Multiple Access (OFDMA) scheme (1024 subcarriers). OFDMA is the technique of choice to allow for flexible bandwidth allocation and multiple access. Frequencies of operation include 2, 3, 5, 7, 8, 10, and 20 GHz in order to allow for flexibility since the standard is desired to be

applied worldwide. The data rates depending on modulation can be up to 63 Mbps, and quality of service can be provided using Multi Protocol Label Switching (MPLS) and Differentiated Service (DiffServ). The network latency is less than 50 milliseconds in order to allow interoperability with 3G cellular networks and the use of VoIP. Time Division Duplex (TDD) and Frequency Division Duplex (FDD) operations are supported, but TDD dominates the deployments. The advantages of TDD are efficient support for downlink and uplink bandwidth, channel reciprocity, and ease of implementation since only one channel is used [42].

The IEEE 802.16 standard has a centralized approach using a base station to fully control the MAC layer. BPSK, QPSK, 16-QAM, 64-QAM and 256-QAM are supported as well as convolutional and turbo codes with variable code rates. Handoff is supported and many security features like a key management protocol and traffic encryption are included to increase the security of the protocol. Multicast and broadcast services are supported and the use of smart antennas is introduced in some of the standard's specifications [42].

The IEEE 802.16 based systems can be used as backhaul links to interconnect the IEEE 802.11-based LANs. The concept can be seen in Figure 21. Two or more 802.11 LANs interconnected using a 802.16 link can provide network connectivity over wide areas. In this scheme, the 802.16 is used as a point-to-point link. The advantage of using 802.16 instead of microwave links is that using one 802.16 base station, tens of 802.11 LANs can be connected at high connection speeds [42].

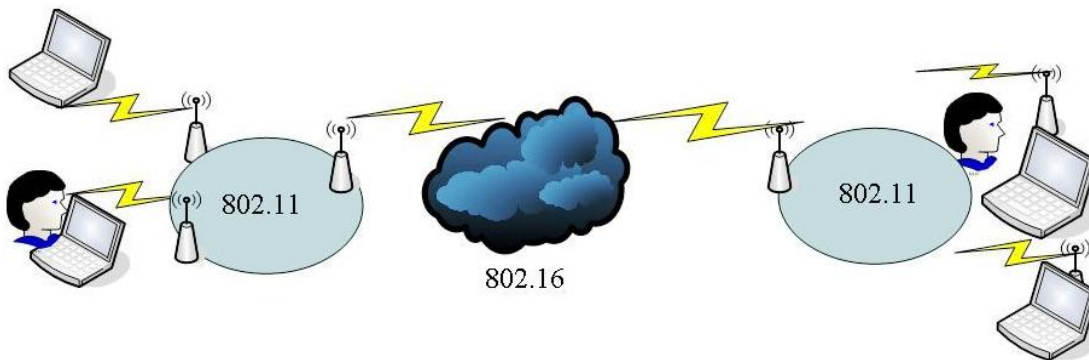


Figure 21. Interconnection of the IEEE 802.11-based LANs using the 802.16 as a Point-to-point Link

E. SUMMARY

In this chapter, digital data transmission has been briefly discussed with an emphasis on the wireless channel. Next, the widely used wireless networking standards are discussed. Modulation and channel coding techniques were briefly described.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. RESULTS

The results reported in this thesis are obtained through Matlab simulation, which is used to investigate the effects of the wireless channel discussed in Chapter III, and experimentation on two commercial VoIP networks. A schematic diagram of the model used to implement the simulation and experiments can be seen in Figure 22. Speech is transmitted through a network which includes at least one wireless link along the end to end path. The speech is received at the other end where the speech recognition procedure takes place.

Two metrics are used to measure the effectiveness of the simulation and experiments. The first is the number of bit errors in the received speech signal compared to the number of bits in the original transmitted signal (bit error rate):

$$BER = \frac{\text{number of bit errors in the received signal}}{\text{number of bits in the original transmitted signal}} \quad (4.1)$$

The second is the amount of comprehensible speech that is received in the receiver (remaining speech) after losses due to packet errors.

In order to measure the amount of remaining speech, speech recognition software is used. First a speech sample, which is used as reference is passed through the speech recognition software, which recognizes all the words of the speech sample and produces a text file. The reference speech sample is transmitted through a wireless channel, which causes distortion to the speech sample due to packet errors. The speech sample that reaches receiver is then applied to the speech recognition software, which typically recognizes a smaller amount of speech since the speech is now distorted. The amount of recognized words in the received speech sample is compared to the amount of recognized words in the original speech sample. The remaining speech is defined as the ratio of the number of words recognized in the received speech to the number of words recognized in the original speech:

$$\text{Remaining speech} = \frac{\text{Number of words Recognized in the Received Speech Signal}}{\text{Number of words Recognized in the original Speech Signal}} \quad (4.2)$$

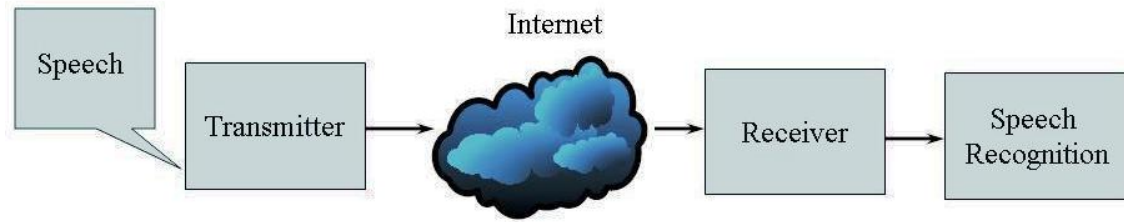


Figure 22. Overall Setup Used for Simulation/experiment Voice Transmission over a Packet Switched Wireless Communication System

A. MATLAB SIMULATION

Matlab was used in order to simulate a wireless VoIP network. The premise of this setup is to simulate the concept of VoIP described in Figure 1. The speech is digitized, compressed, packetized, and transmitted; the receiver then follows the reverse procedure.

In order to implement the setup, Matlab, Speex and Dragon Naturally Speaking software packages were used. Speex is a voice compression software package based upon the CELP technique. Some details of the software can be seen in Appendix A. Dragon naturally Speaking is a commercially available voice recognition software package. A short description of Dragon Naturally Speaking is provided in the Appendix A.

A speech recording is first input to Speex. After the speech is compressed, it is exported to Matlab, which simulates the various fading channels. After passing through the simulated channel, the received signal is decompressed using Speex. The decompressed signal is input to the speech recognition software, which typically recognizes only a part of the speech sample depending on the distortion applied to the speech signal in the fading channel. The amount of recognized words in the distorted speech sample is compared to the amount of recognized words in the original speech sample. The overall implementation is shown in Figure 23.

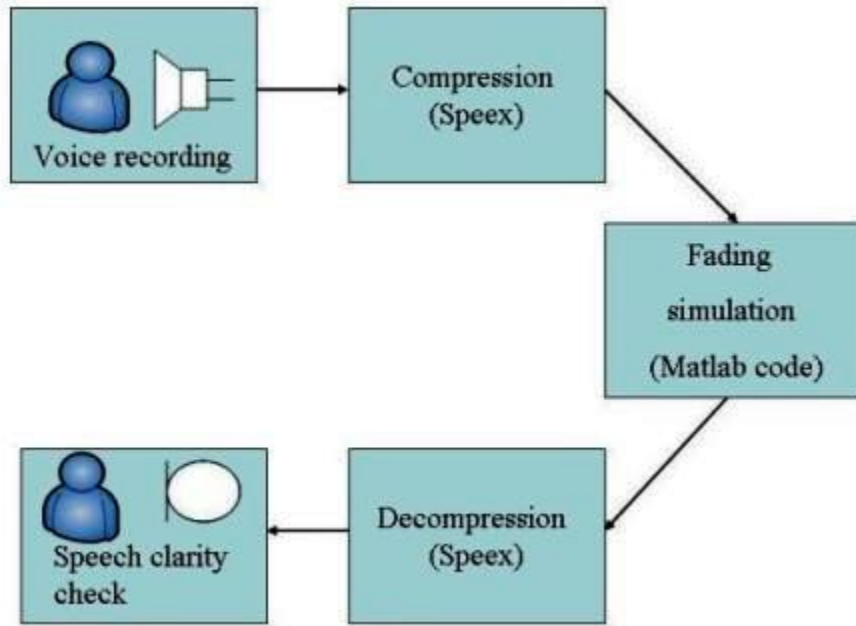


Figure 23. Matlab and Additional Simulation Software Setup used to Simulate VoIP over a Wireless Network

Four wireless channels were implemented for the needs of this simulation: Rician and Rayleigh fading channels in additive white Gaussian noise (AWGN) and Rician and Rayleigh channels in AWGN with convolutional coding. Additionally, a Matlab simulation was implemented to simulate a fading channel for an audio file without compression for the purpose of comparing results of the simulation. The Matlab code is included in Appendix B.

1. Rician Fading Channel in AWGN

In this implementation, the compressed speech samples are modulated (baseband) and transmitted through a Rician fading channel in additive white Gaussian noise. After demodulation and decompression in the receiver, the BER is determined. The K factor of the Rician channel was set to one and the noise level could be fixed or variable. A special case of this simulation is when the K factor was set to zero, where the channel becomes Rayleigh.

2. Rician Fading Channel in AWGN and Convolutional Coding

A convolutional coder with $(n, k, K) = (2, 1, 3)$ is added before the signal is modulated. In the receiver, after the signal is demodulated, the coded bits are decoded to extract the received speech bits. The K factor of the Rician channel was again set to one, and the noise level could be fixed or variable.

3. Simulation Results

This section presents the results obtained from the Matlab-Speex simulation.

a. Rician Fading Channel: K factor

The first simulation examined the effect of varying the K factor of a Rician channel on bit error rate. The results of the simulation are plotted in Figure 24 and were obtained based on averaging results from 50 Monte Carlo runs. By increasing the K factor, the BER decreases from a value close to 0.06 for $K = 0$ down to 10^{-5} for a K factor of 25. For $K = 0$, the channel becomes a Rayleigh channel, and, for $K = \infty$, it is an additive white Gaussian noise channel. The case of $K = 0$ represents the worst case scenario, which yields a high BER that renders it impractical for transmission of information. For an increase of the K factor of an order of magnitude, the improvement in BER is between one and two orders of magnitude.

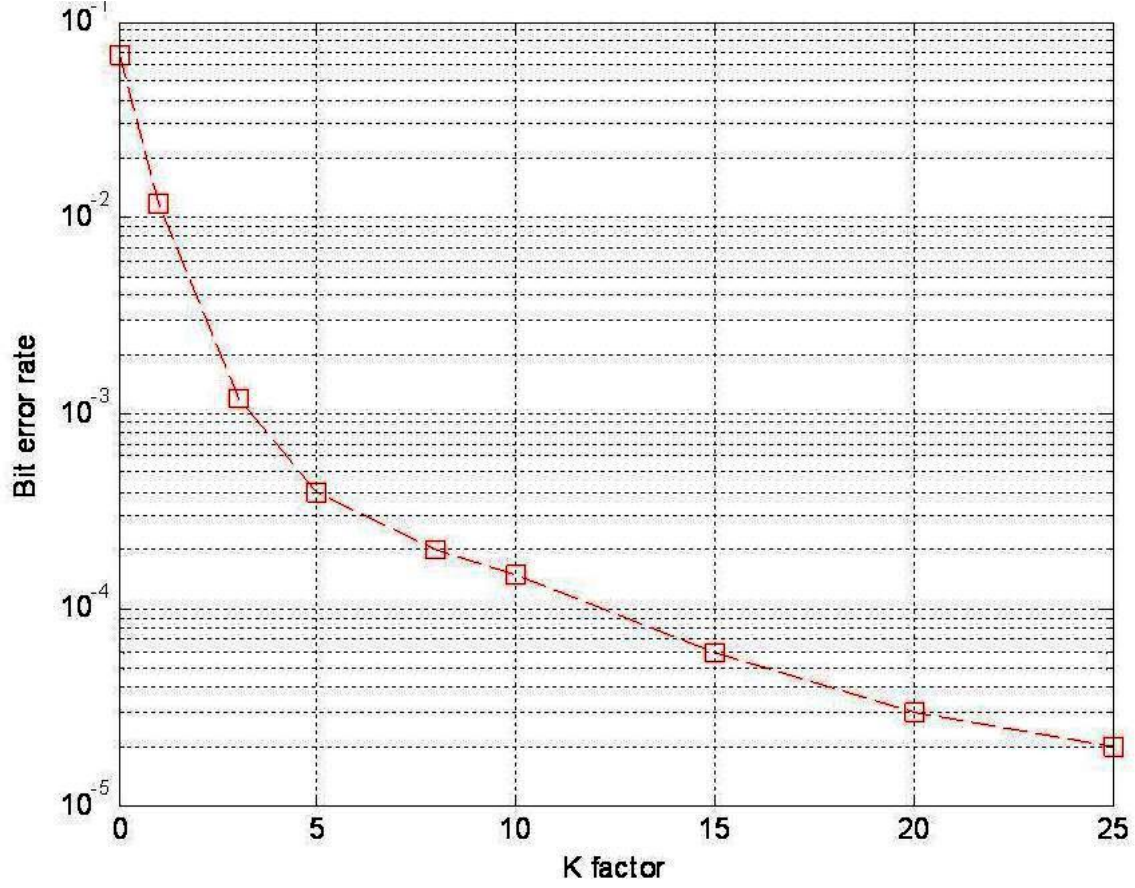


Figure 24. The Bit Error Rate as a function of the Ratio of Dominant over Secondary Path (K factor) for the Rician Fading Channel based on 50 Monte Carlo Simulation Runs

b. Rician Fading Channel: SNR

The next is the simulation of a Rician fading channel without convolutional coding to study the effects of SNR on the bit error rate. All the parameters of the channel remain constant except for the signal- to-noise ratio (SNR) of the dominant path. SNR in dB is defined as:

$$\text{SNR}_{dB} = 10 \log_{10} \frac{P_s}{\sigma_n^2} \quad (4.3)$$

where P_s is signal power and σ_n^2 is the noise variance. The results are plotted in Figure 25. At SNR = 6 dB, the BER = 0.5, which makes the channel inappropriate for transmission of information. As the SNR increases from 6 dB to 23 dB, the BER decreases and reaches values close to 10^{-6} . For values of SNR between 6 dB and 16 dB,

there is no significant improvement in the BER. For values of SNR of more than 16 dB, there is a rapid improvement in the BER. An improvement of about one order of magnitude is obtained for an increase in SNR from 22 dB to 23 dB. The general remark for this simulation is that, by increasing the SNR of the dominant path, the BER of the transmission decreases.

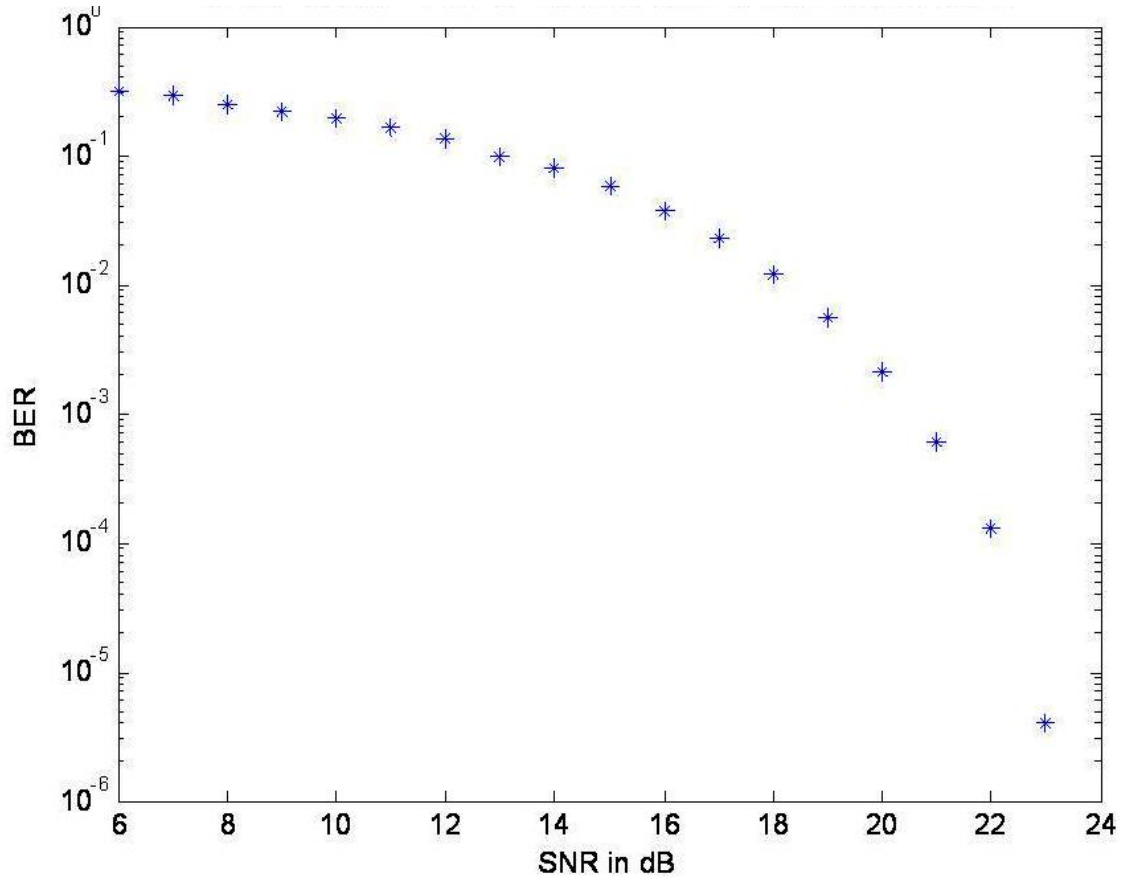


Figure 25. Effect of SNR on Bit Error Rate for a Rician Fading Channel based on 50 Monte Carlo Simulation Runs

c. *Rician Fading Channel With Convolutional Coding*

After measuring the effects of varying the SNR of the dominant path in a Rician fading channel, the same channel was used to determine the effects on BER when channel coding is used. Furthermore, the effects of transmitting compressed and uncompressed speech signal through the same channel are also examined. The results are shown in Figure 26 and were based on 50 Monte Carlo simulation runs.

Comparison the BER plots of the compressed and uncompressed signal through the same channel, there is no significant difference between the two cases. What makes a difference is the amount of audible distortion caused in each case for the same amount of errors. Specifically, the uncompressed signal with a BER of 10^{-3} presents a noticeable amount of distortion but it is still understandable. On the other hand, the compressed signal is not even decodable. It is easy to realize that compressing a speech signal results in a gain in bit rate, but the signal becomes more sensitive to errors.

For all the cases shown in Figure 26, an increase in SNR leads to a decrease of BER regardless of channel coding or speech compression. This effect is due to multipath. As the dominant path becomes stronger, the uncertainty about ISI and consecutive pulse discrimination decreases. When the strength of the main path becomes strong enough, it is easier for the receiver to discriminate between a pulse and a delayed copy of a previous pulse.

Comparing the uncompressed and compressed speech, as the SNR of the dominant path increases, it is seen that, after a threshold value of SNR, there is a coding gain increase as the SNR increases. More specifically, after a SNR of 13 dB where there is no coding gain, an increase in coding gain is obtained. A maximum coding gain of 5 dB is obtained for $\text{BER} = 10^{-5}$. Furthermore, one may notice that, before the threshold value of $\text{SNR} = 14$ dB, there is a negative coding gain, meaning that the results are better without channel coding rather than with it.

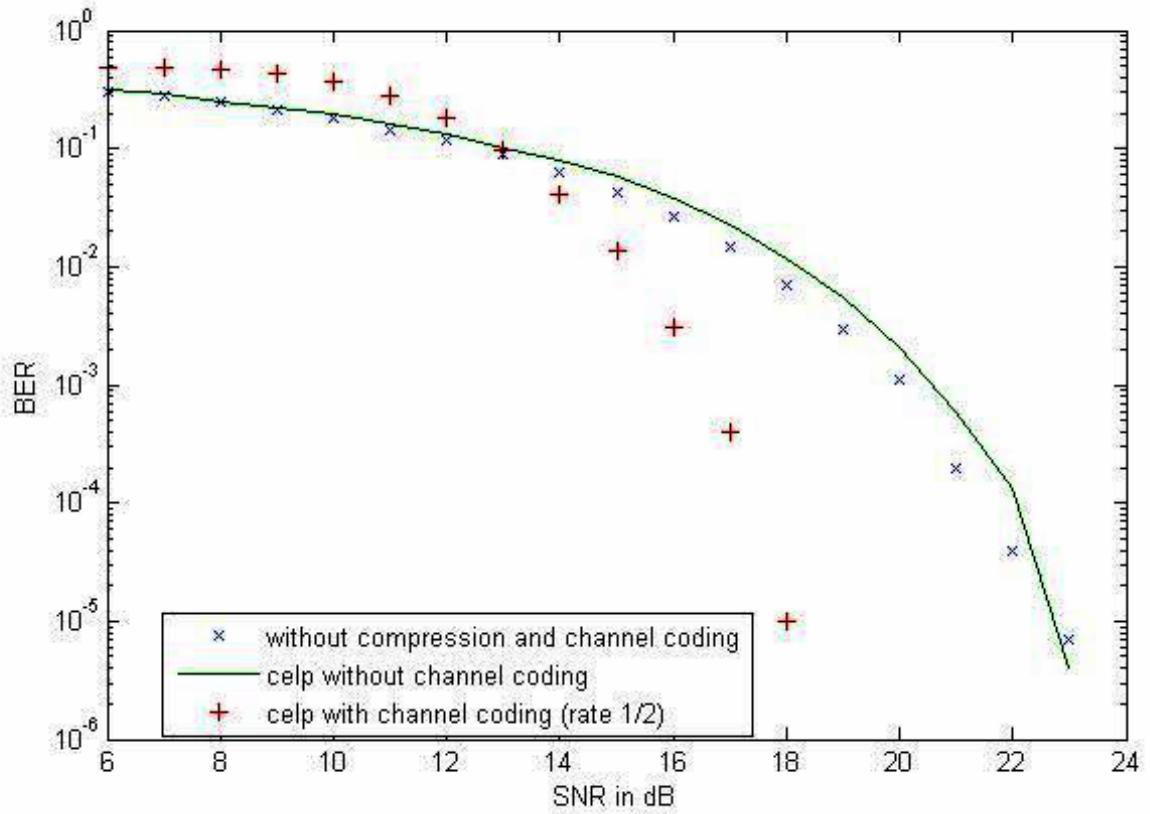


Figure 26. Effects of SNR on the BER for a Rician Fading Channel with Convolutional Coding, and Speech Coding based on 50 Monte Carlo Simulation Runs

d. Secondary Path Delays in a Rician Fading Channel

The Rician fading channel is next tested for the effect of secondary path delay variation on the BER of the signal. The results can be seen in Figure 27 and were based on 50 Monte Carlo simulation runs.

An increase in the secondary path delay variation causes an increase in the BER of the signal. After a 30 ns delay is inserted, it is noticed that the BER reaches a value close to 0.5. The reason for this result comes immediately from the effect of multipath. As the signal strength in the paths that the secondary signals follow become larger, the ISI distortion increases. When the delay variation of the secondary path becomes too large, the receiver is unable to discriminate between a pulse and the delayed copy of a previous pulse.

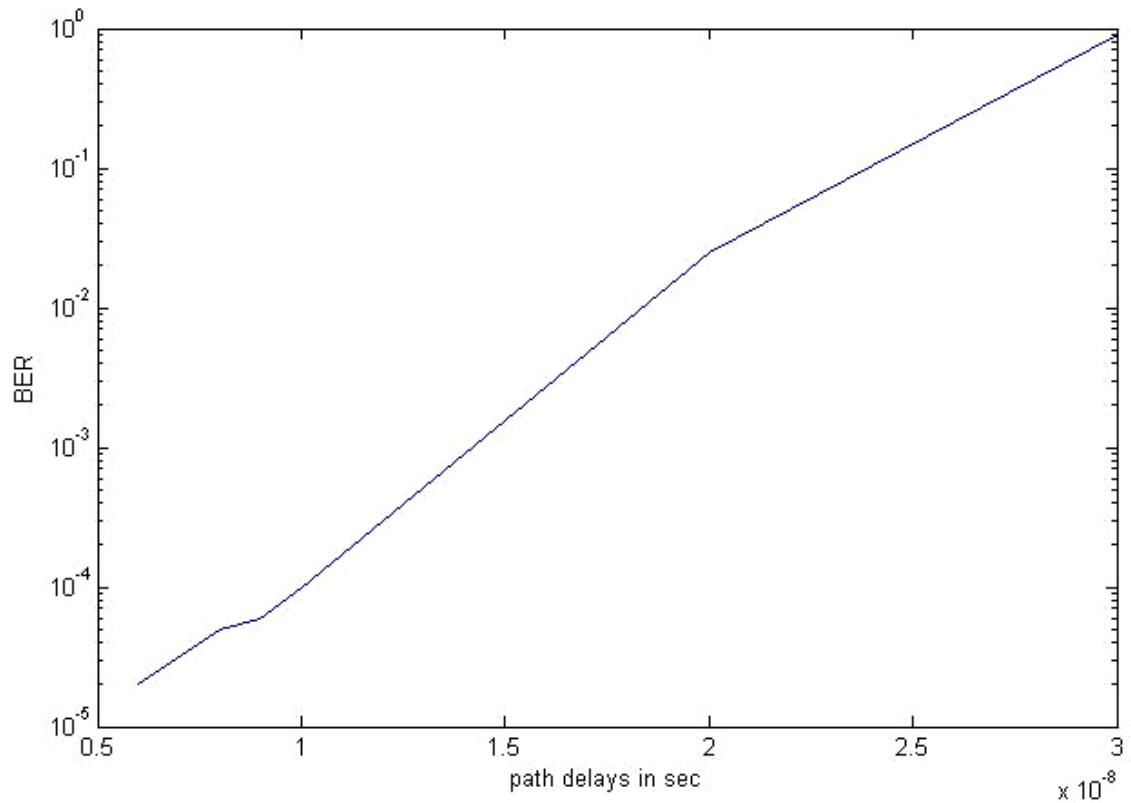


Figure 27. Effect of Secondary Path Delays on the BER of a Rician Fading Channel based on 50 Monte Carlo Simulation Runs

e. Effect of Secondary Path Signal Strength in a Rician Fading Channel

The effect of the secondary path signal strength on the BER of a Rician fading channel is also examined. The results are shown in Figure 28 and were based on 50 Monte Carlo simulation runs.

As the signal strength of the secondary paths increases, the BER increases as well. This result is logical if one considers that as the secondary signals get stronger, they make the discrimination of a pulse and a delayed copy of a previous pulse a harder task for the receiver. For the specific channel, when the secondary paths are -7 dB weaker than the main path signal (which is 0 dB), it is impossible for the receiver to correctly detect the pulses (BER reaches 0.5). On the other hand, when the secondary paths are -15 dB weaker than the main path signal (which is 0 dB), the BER is 10^{-6} .

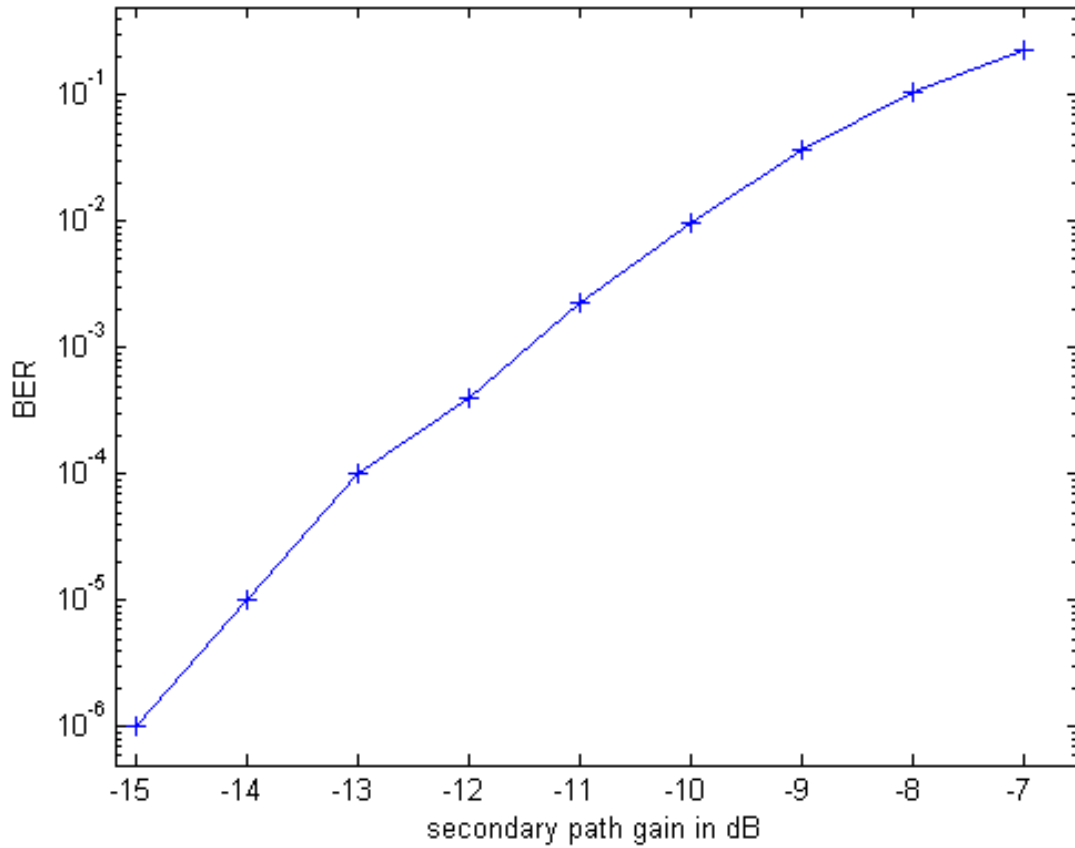


Figure 28. BER as a Function of Secondary Path Gain for a Rician Fading Channel based on 50 Monte Carlo Simulation Runs

f. Remaining Speech after Decompression as a Function of Compression Ratio

Next, the effect of compression ratio on the speech quality is examined. In order to determine the results of these measurements the setup of Figure 23 was used. Five different compression ratios were used, and the results of the simulation can be seen in Figure 29. Sixty Monte Carlo runs were used to calculate the average amount of remaining speech for each compression ratio.

As the signal is compressed at higher rates, the amount of remaining speech becomes smaller. It is expected that the more compressed a signal is, the more “sensitive” it is to the effects of errors. Every bit in a compressed signal represents a

larger amount of data than in an uncompressed signal. Thus, when losing a bit that represents compressed data, the amount of information lost is much more than in the uncompressed case. For the simulation under discussion, by decreasing the compression ratio from 10:1 to 2:1, there is an increase in the amount of remaining speech from 0.01 to 0.26 of the original speech sample.

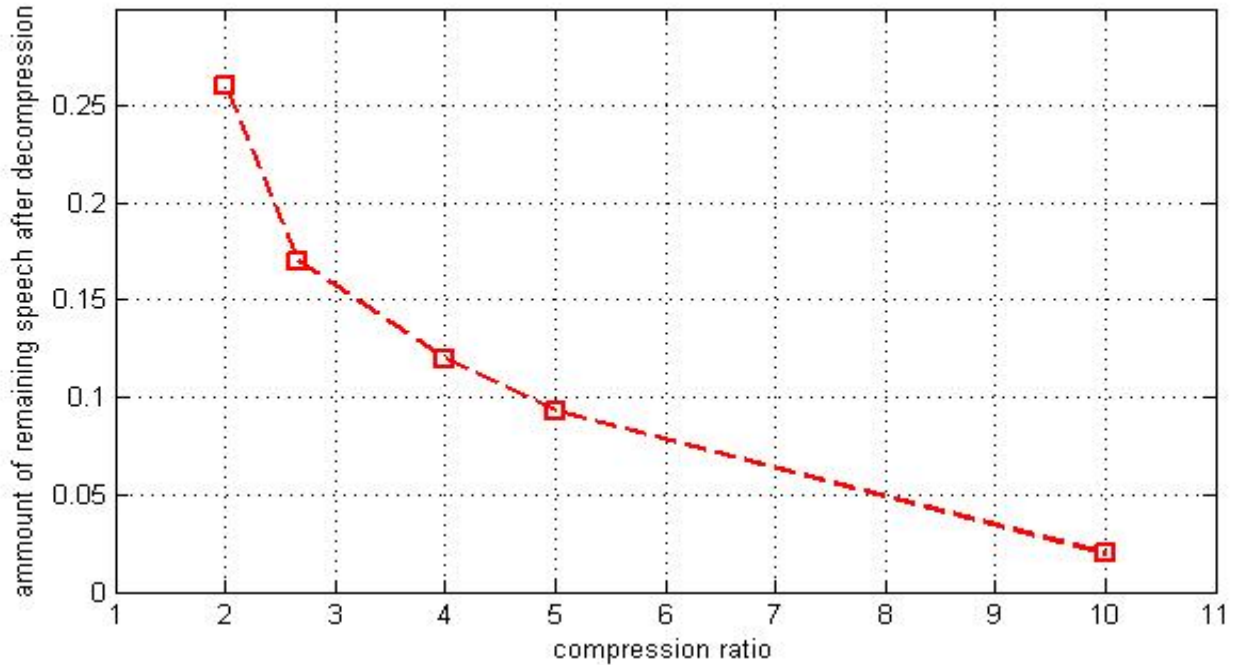


Figure 29. Effect of Compression Ratio on the Remaining Speech after Decompression based on 50 Monte Carlo Simulation Runs

g. Remaining Speech after Decompression as a function of Compression Ratio With and Without Channel Coding

After determining the effects of compression on the amount of received speech, channel coding was introduced to determine its effects on the amount of remaining speech. The same setup as in the previous subsection was used, and a simulation of 50 Monte Carlo runs was executed. Two different rates were used for the convolutional coding, $\frac{1}{2}$ and $\frac{3}{4}$, and both gave the same results, which can be seen in Figure 29. Similar to the previous results, the amount of remaining speech decreases as

the signal compression ratio is increased. When channel coding is used, not only the remaining speech quality improved, but also the errors were eliminated in comparison to the case of without channel coding.

Another important conclusion comes from comparing the amounts of compression and their results with and without coding. It is preferable to use a high compression ratio (10:1) with convolutional coding rather than use a lower compression ratio (2:1) without convolutional coding. If a 10:1 compression ratio with a channel coding rate of $\frac{1}{2}$ is used, a total of 6 kbps is transmitted, but 100% the speech is received. On the other hand, by using a 2:1 compression ratio without channel coding, the result is a total transmitted signal of 40 kbps, but the received speech is only 30% of the original speech. The drawback of channel coding is circuit complexity, cost, and delay, which are important in real-time applications.

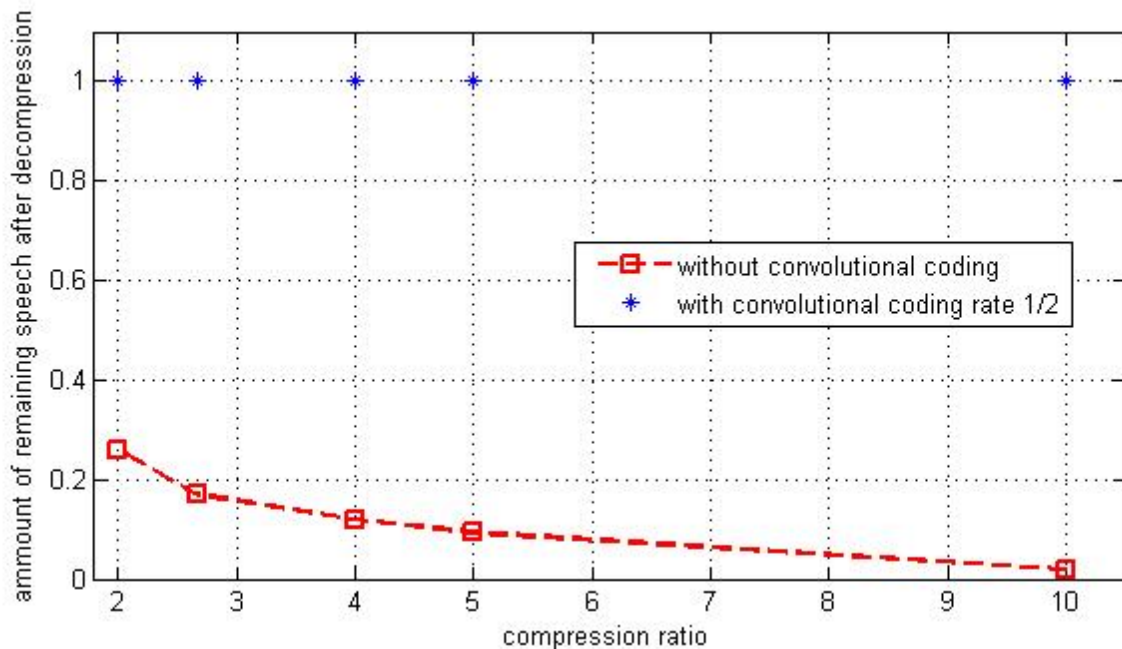


Figure 30. Effect of Compression Ratio on the Remaining Speech after Decompression with Channel Coding based on 50 Monte Carlo Simulation Runs

B. VOIP EXPERIMENTS OVER A WIRELESS LAN

We now examine the voice quality in terms of packet loss and delay in commercial VoIP networks. Experiments were conducted on two different platforms, namely Skype and Vonage.

1. Skype Service

The implementation used to conduct the VoIP experiments in the Skype network can be seen in Figure 30. Two users are connected to a LAN, one with wireless access and with the other with wired access. During the call setup phase, the two users communicate through Skype's servers for signaling purposes. Once the call setup is completed, the two users can communicate directly and are one hop away (since they are separated only by a bridge within the same LAN). The measured transmission delay during data transfer (using ping and traceroute) was less than 1 ms. After the two users are connected and can communicate, User 1 transmits a recorded signal and User 2 records it. The signal travels from User 1 to User 2 only through the router, which is verified by using a packet sniffer and checking the TTL value of the received packets.

By changing the position of the wireless user relative to the access point, attenuation and fading are inserted into the communication path, and its effects are studied. After the received speech is received, it is passed through the speech recognition software in order to measure the amount of words recognized from the software. This amount of words is smaller than the amount of words recognized on the original speech sample. Comparing the amount of words recognized on the original and final speech sample, we define the percentage of remaining speech as a measure of degradation in quality.

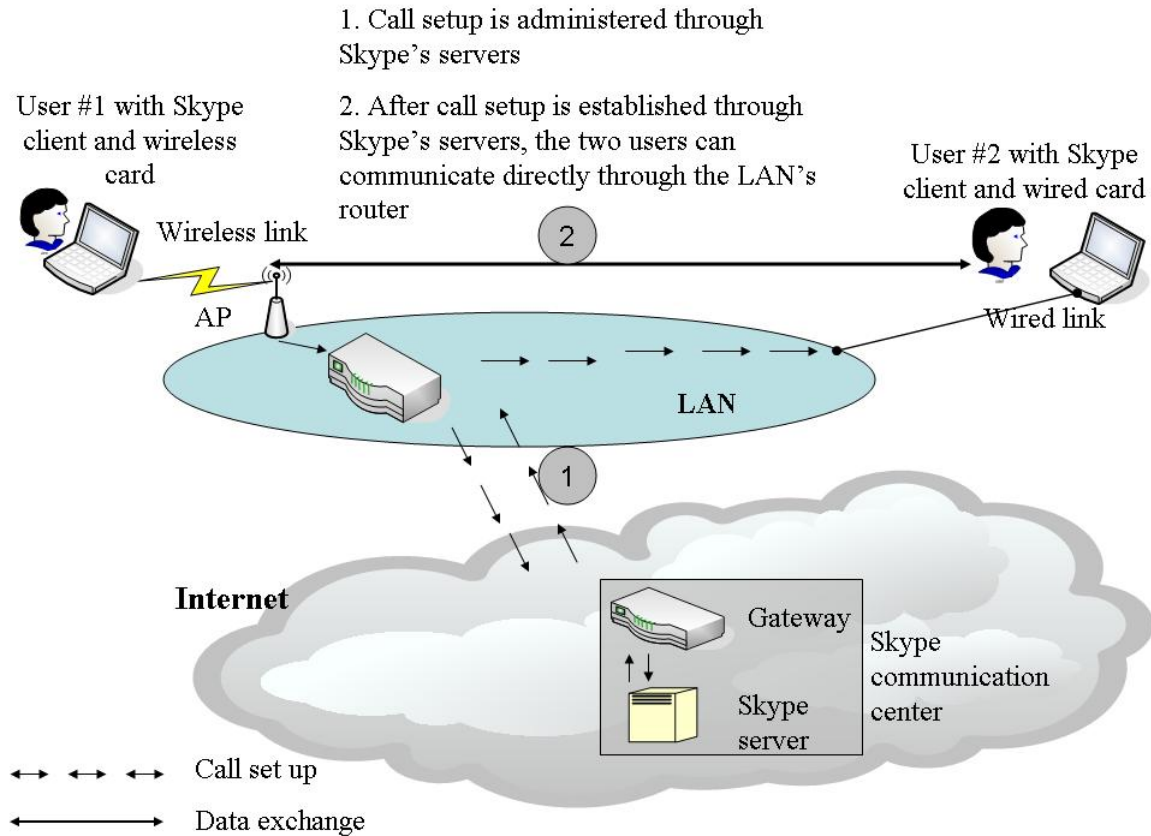


Figure 31. Skype Implementation with Two Users Interconnected with VoIP Inside the Same LAN

Two sets of measurements were obtained, one for the indoor and one for the outdoor environment. In both cases, the average receiver signal strength (in dBm) was measured for every receiver position along with the remaining speech. The receiver's signal strength was measured using two different packet sniffers simultaneously (Cain and Ethereal). For every position of the receiver, the average signal strength at ten-minute periods was recorded. The average measurements were based on twelve repetitions of the experiment. The results can be seen in Figure 32.

For the outdoor environment, 100% of the speech signal is received up to a signal strength of -80 dBm. Below -80 dBm, there is a rapid degradation of the remaining speech. At -85 dBm, seven out of twelve times, the Skype client was logged off the network. Furthermore, two out of twelve times, the Internet connection was lost, and the laptop had to be reconnected to the wireless network. For the case of -90 dBm, no connection could be established between the laptop and the wireless LAN.

For the indoor environment, the whole speech signal is received up to a signal strength of -75 dBm. Below -75 dBm, there is a rapid degradation of remaining speech. At -78 dBm, three out of twelve times the Skype client was logged off the network, and one out of twelve times the Internet connection was lost, and the laptop had to be reconnected to the wireless network. At -80 dBm, eight out of twelve times the Skype client was logged off the network. For the case of -90 dBm, no connection could be established between the laptop and the wireless LAN.

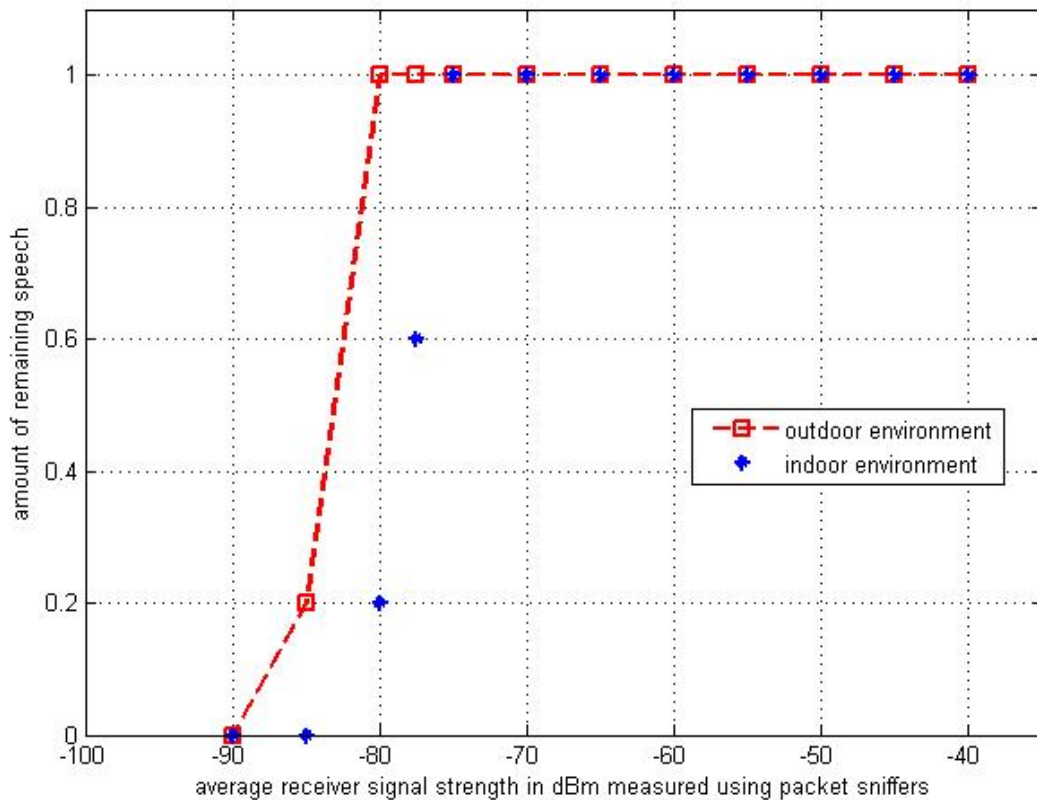


Figure 32. Amount of Remaining Speech versus Average Receiver Signal Strength for Skype Measurement Setup based on 12 Monte Carlo Runs

Comparing the results of the indoor and outdoor environments, it can be concluded that the wireless network performs better in the outdoor case. More specifically, for the outdoor scenario, there are no significant obstructions, and the transmitted signal suffers mainly from attenuation. There is one direct path and the multipath effect is limited. On the other hand, for the indoor scenario, there are

significant obstructions (walls, furniture, etc.) causing the multipath effect. As a consequence, the transmitted signal suffers from fading and attenuation, which in turn leads to degradation compared to the outdoor case.

2. Vonage Service

The implementation used to conduct the VoIP experiments using Vonage soft phones can be seen in Figure 29 and was used as an alternative platform to Skype. Two users are connected to a LAN using a wireless and wired connection as in the Skype implementation. During the call setup phase, the two users communicate through Vonage servers using SIP. Once the call setup is completed, the two users continue to communicate through the servers. The network architecture is totally different from the one used in Skype. The two users are ten hops away since they are not only separated by a router but also the data packets travel up to New Jersey and back where the Vonage servers are located. The measured transmission delay (measured using traceroute and VisualRoute) was on average 30 ms. The signal travels from User 1 through the router and to the Vonage servers in New Jersey through the Internet and then back to the LAN's router and finally to User 2. The traveled route was verified by using a packet sniffer and checking the TTL value of the received packets.

After the call setup is established User 1 transmits a recorded signal and User 2 records it. Attenuation and fading are inserted the same way as in the Skype implementation, and the speech recognition software is used to measure remaining speech.

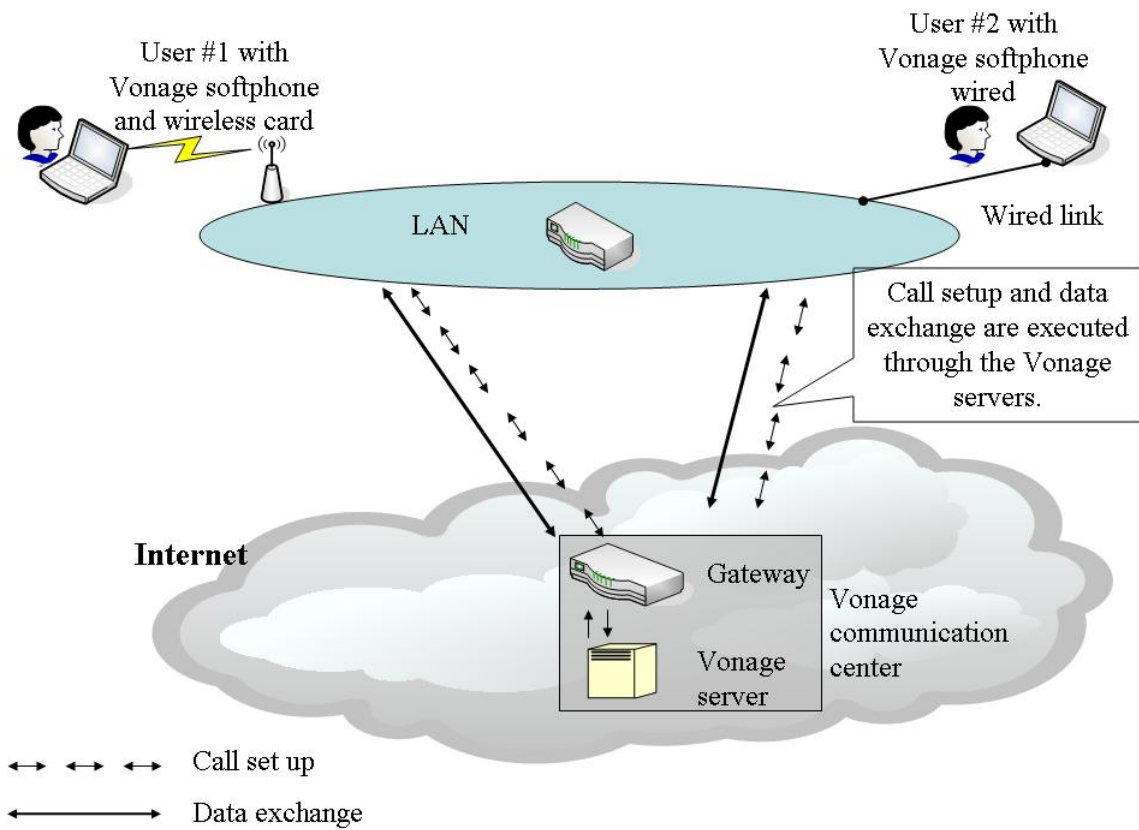


Figure 33. Vonage Implementation as an Alternative Platform to Skype. The Two Users are in the Same LAN but the Signal Travels through Vonage Servers in New Jersey during Both the Call Setup and the Data Exchange Phase

Two sets of experiments were conducted as in the case of Skype, one for the indoor and one for the outdoor environment. In both cases, the average receiver signal power (in dBm) was measured for every receiver position along with the remaining speech. The average measurements were based on twelve repetitions of the experiment. The results can be seen in Figure 34.

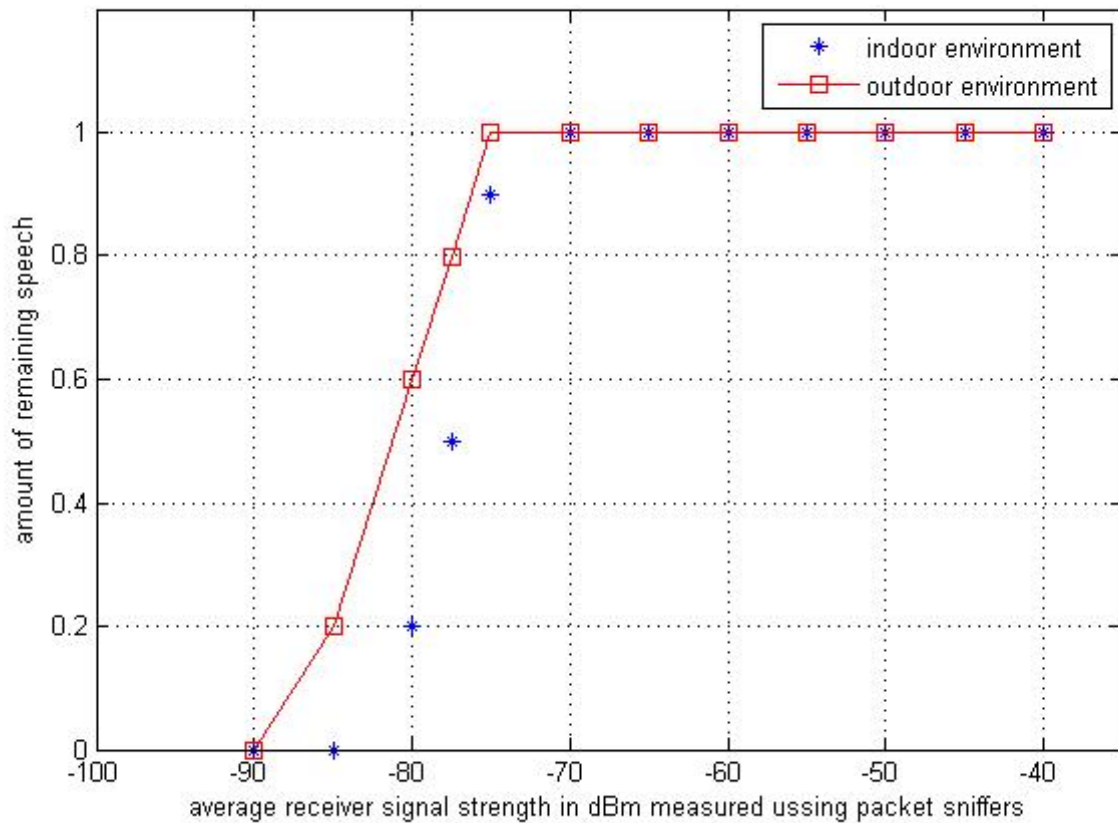


Figure 34. Amount of Remaining Speech versus Average Receiver Signal Strength for Vonage Experimentation

For the outdoor environment 100% of the speech signal is received up to a signal strength of -75 dBm. Below -75 dBm, there is a rapid degradation of the percentage of remaining speech and two out of twelve times the Internet connection was lost, and the laptop had to be reconnected to the wireless network. For the case of -90dBm no connection could be established between the laptop and the wireless LAN.

For the indoor environment the whole speech signal is received up to a signal strength of -70 dBm. Below -70 dBm, there is a rapid degradation of the remaining speech. At -85 dBm, four out of twelve times the Internet connection was lost and the laptop had to be reconnected to the wireless network. At -90 dBm, no connection could be established between the laptop and the wireless LAN. For both the outdoor and indoor environments, no client log off was observed.

Comparing the results of the indoor and outdoor environments, it can be concluded that, similar to the Skype case, the wireless network performs better in the outdoor case. This is, again, due to the fading effect occurring mainly at the indoor environment.

Comparing the results for Skype and Vonage, it is noticed that Skype achieves a slightly better performance for both the outdoor and indoor environment. This is mainly due to the delay inserted in the Vonage implementation when the signal travels during the data transfer phase from Monterey to New Jersey (about 30 ms). In addition to the longer delay, the path loss and multiple hops contribute to a higher packet loss in the Vonage case.

C. VOIP EXPERIMENT ON A WAN

A VoIP experiment was conducted on a WAN network to examine the effectiveness of VoIP during a 24-hour period on a long-distance connection. The experiment was conducted on the Skype platform.



Figure 35. VoIP Measurements on a WAN Implementation. The Two Users are Located in Monterey, California and Athens, Greece, Respectively. One User Transmits a Recorded Message through VoIP Using Skype and the Other Records it.

The setup of the experiment can be seen in Figure 35. One user is located in Monterey, USA and the other in Athens, Greece. The two users are located on average 12 hops away and the average transmission delay is measured (using tracert) to be between 200 ms and 360 ms depending on the time of the day. The physical distance between the two users is about 7000 miles. Two main paths are followed depending on the time of the

We remark that the measurements seem to follow a random pattern. There are time periods during which the quality of communication is higher, but one cannot predict a time slot with guaranteed quality. This is due to the dynamic nature of the Internet traffic. The path followed each time varies and with it, the delay and packet loss vary as well.

During the late evening hours (from 1700 to 0300 hours Pacific Time) there is degradation in the quality of speech. During these hours, it is the peak usage time in Europe's networks, which creates a bottleneck for congestion in the implementation. During these hours, the wired lines between East Coast and Europe were observed to be congested and the signal traveled through slower lines (e.g., satellite) causing additional delay and thus a decrease in speech quality.

Of the 288 recordings, only 20 were completely recognized by the recognition software, giving an approximately 7 % rate of success in complete recognition. For the majority of the cases, the quality of the received speech was satisfactory, with a remaining speech value of more than 0.95. Fifteen of the 288 recordings yielded a remaining speech value of less than 0.90, and the worst case recorded was 0.72.

In conclusion, the speech quality on a long-distance communication is affected by the time of the day. In the reported experiment, there was degradation in speech quality during late evening hours because of the rush hour in Europe's networks.

D. SUMMARY

In this chapter, the implementations used to examine and simulate a VoIP network were presented. Two implementations were used. The first was Matlab-based, which required the use of Speex for speech compression, and examined the effects of wireless channel, compression ratio and recognition quality of the received speech. The second consisted of experiments on two commercial VoIP networks in order to measure speech recognition and comprehension. The results of these simulations and experiments were reported.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS

This thesis investigated the quality of received voice with emphasis on the effects of wireless channel, speech compression and channel coding. Matlab, Speex, and Dragon Naturally Speaking software were used to simulate VoIP communication. Matlab was used to simulate various wireless channels and Speex was used to compress speech signals using CELP. The simulation quantified the effects of wireless channel, compression ratio and channel coding on the received speech quality. The metrics used were the BER and the amount of speech that remained at the receiver's end. The wireless channels simulated were Rician fading channels with additive white Gaussian noise and convolutional coding.

Next, the voice quality in terms of packet loss and delay in commercial VoIP networks was examined using experimentation. Experiments were conducted on two different platforms, namely Skype and Vonage. The first experiment used a LAN, and investigated the effects of architectures of the two providers on the received speech quality. Finally, VoIP measurements were made on a WAN network to examine the effectiveness of VoIP during a 24-hour period on a long-distance connection. The experiment was conducted using the Skype platform.

A. SIGNIFICANT RESULTS

Simulations showed that for the Rician fading channel, an increase in the SNR causes a decrease in BER. There is no significant difference between the BER of the signal when transmitting compressed and uncompressed speech. What makes a difference is the amount of audible distortion caused in each case for the same amount of errors. The increase in the secondary path delay variation causes an increase in the BER of the signal. As the signal strength of the secondary paths increases, the BER increases as well.

Both Skype and Vonage experiments showed a fast degradation of the percentage of remaining speech after a threshold signal strength value. Performance of an outdoor wireless network was better than that of an indoor network due to the effect of multipath occurring indoors. Comparing the results for Skype and Vonage, it is noticed that Skype

achieves a slightly better performance for both the outdoor and indoor environment. The architecture of the Vonage network causes additional delay, path loss and multiple hops that contribute to a higher packet loss.

A second experiment used VoIP over a WAN. The results follow a random pattern due to the dynamic nature of the Internet traffic. The path followed each time varied and with it the delay and packet loss. Degradation of speech quality was observed during the rush hours due to network congestion. During these rush hours, the signal had to travel through slower lines (e.g., satellite) causing additional delay and thus a decrease in speech quality.

B. FUTURE WORK

This study was based on simulation in Matlab and experiments on commercial VoIP networks. In both cases, improvements as well as additions can be made.

In this work, simulation was focused on a specific kind of baseband modulation, without investigating the effects of different modulation schemes on the quality of the received speech. It was observed though that different modulation schemes used in a wireless network can affect the network performance and thus the VoIP communication quality. We suggest an investigation, through simulation, of the effects of modulation on the received speech quality of VoIP over wireless communications.

Experiments conducted in this work used VoIP over LAN and VoIP over the Internet. The limitations of limited access to a satellite and no access to an IEEE 802.16 link as part of the overall network did not permit the investigation of their effects on the VoIP communication quality. It is proposed that, in a future effort, a correlation of traceroute paths indicating satellite links to remaining speech would be attempted. Also an IEEE 802.16 link should be included as part of the network in order to investigate the effect of these links on the quality of the received speech.

The experimentation over a long distance network conducted in this thesis was limited to a 24-hour recording period. The results thus acquired indicated a trend but are

not statistically significant due to the small amount of time the experiment lasted. It is proposed to extend the period of experimentation in order to achieve statistically significant results.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A

A. SPEEX OVERVIEW

SPEEX is a multiple bit-rate speech codec that uses CELP as the encoding algorithm. It supports 8, 16, and 32 kHz sampling rates. It is designed for use in VoIP and for that it has built-in robustness to packet loss. The main characteristics are its flexibility and that it is a free software (open-source). The compressed bit rates supported range from 2 to 44 kbps. It uses voice activity detection (VAD) and has variable complexity selected at the time of compression/decompression. It supports both stereo and mono options and it is a fixed-point implementation [43].

The algorithmic delay of Speex depends on the sampling frequency used and is equal to 30 ms for the 8 kHz and 34 ms for the wideband 16 kHz.

Figure 37 illustrates how the Speex software is used in this work.

B. DRAGON NATURALLY SPEAKING

Dragon Naturally Speaking is a commercial software package. It is a voice recognition software from Nuance. It uses continuous speech recognition and requires that the software be installed and then trained by the specific user that will use it. Even though it is not clearly stated in the documentation, it seems that it uses the HMM algorithm for recognition.

It enables the user to dictate a message, instead of typing it to the screen and can be used to write a letter through dictation and then revise it without the use of the mouse and keyboard. It can be used to browse the Web, start programs and work on them, and create custom commands and command scripts by dictation. The recognition performance improves with use since the software is continually being trained while it is being used. Figure 37 illustrates how the Speex software is used in this work.

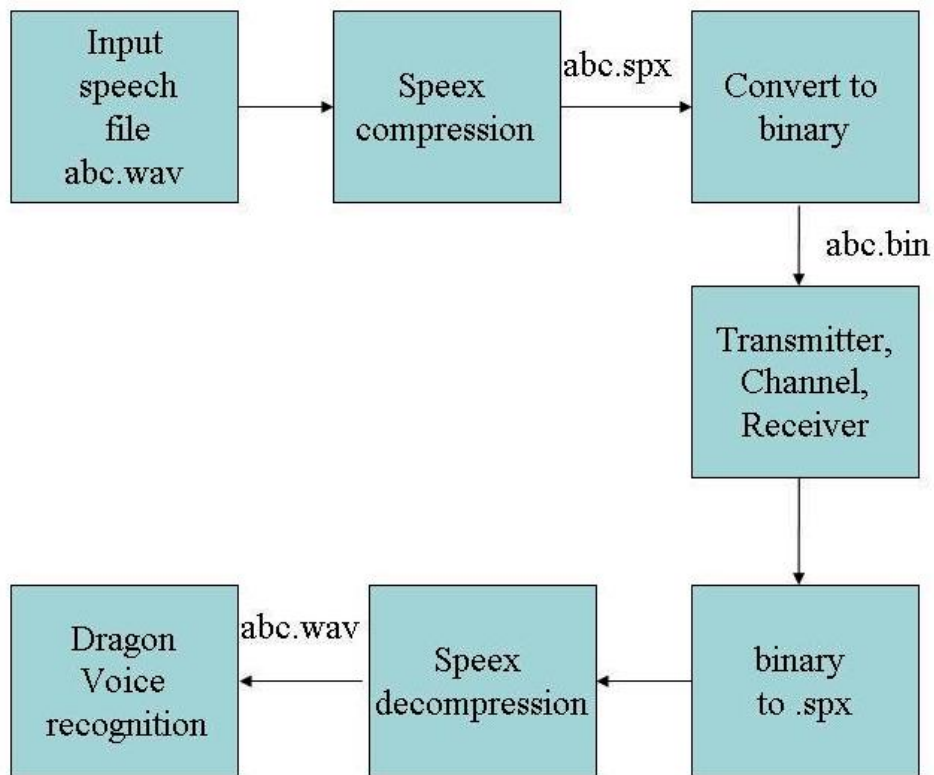


Figure 37. Schematic Diagram of Speex and Dragon Naturally Speaking Interconnection with Matlab

APPENDIX B

MATLAB CODE

This appendix includes the Matlab code developed to conduct the simulation studies reported in this thesis.

```
%%Matlab code to simulate a Rician fading channel with AWGN without
%% convolutional coding on a Speex compressed file
clear
clc
%%open and read the compressed file
fid=fopen('zipar','r+')
c=fread(fid,inf);
fclose(fid)
%%convert to binary
bin=dec2bin(c);
diplo=double(bin);
diplo2=diplo-48;
telbin=reshape(diplo2,1,(27540*8));
telbin=telbin';
%create the Rician channel
chan = ricianchan(1e-8,300,1);
%take delay into account
delay = chan.ChannelFilterDelay;
M = 2;%modulation factor
pskSig = dpskmod(telbin,M);
%insert the effects of channel into signal and add noise
fadedSig = filter(chan,pskSig); rxsig=awgn(fadedSig,10);
%receive the signal
rx = dpskdemod(rxsig,M);
```

```

tx = telbin(2:end); rx1 = rx(2:end);
tx_trunc = tx(1:end-delay); rx_trunc = rx1(delay+1:end);
%measure BER
[num,ber] = biterr(tx_trunc,rx_trunc)
a=zeros(1);
rx=rx';
%reshape the signal to bring it to its original shape
xanabin=reshape(rx,27540,8);
xanabin=xanabin+48;
kordoni=char(xanabin);
dek=bin2dec(kordoni);
leles=double(dek);
save neo1.dat leles

```

```

%%Matlab code to simulate a Rayleigh fading channel with AWGN and without
%% convolutional coding on a Speex compressed file
clear
clc
%save memory space
cwd = pwd;
cd(tempdir);
pack
cd(cwd)
%open the compressed file
fid=fopen('zipar','r+')
c=fread(fid,inf);
fclose(fid)
bin=dec2bin(c);
diplo=double(bin);
diplo2=diplo-48;
telbin=reshape(diplo2,1,(27540*8));
telbin=telbin';

```

```

%create the Rayleigh channel and account for delay
chan = rayleighchan(1e-8,4,[0 1e-7],[0 -14]);
delay = chan.ChannelFilterDelay;
M = 2; %modulation order
pskSig = dpskmod(code,M);
%pass the signal through channel
fadedSig = filter(chan,pskSig);
rxsig=awgn(fadedSig,61);
%demodulate
rx = dpskdemod(rxsig,M);
tx = telbin(2:end); rx1 = rx(2:end);
tx_trunc = tx(1:end-delay); rx_trunc = rx1(delay+1:end);
%find BER
[num,ber] = biterr(tx_trunc,rx_trunc) % Bit error rate
a=zeros(1);
rx=rx';
%bring signal to original shape
xanabin=reshape(rx,27540,8);
xanabin=xanabin+48;
kordoni=char(xanabin);
dek=bin2dec(kordoni);
leles=double(dek);
save cal30last40krun12.dat leles

```

```

%%Matlab code to simulate a Rayleigh fading channel with AWGN and
%% convolutional coding on a Speex compressed file
clear
clc
cwd = pwd;
cd(tempdir);
pack
cd(cwd)

```

```

fid=fopen('zipar','r+')
c=fread(fid,inf);
fclose(fid)
bin=dec2bin(c);
diplo=double(bin);
diplo2=diplo-48;
telbin=reshape(diplo2,1,(27540*8));
telbin=telbin';
%convolutional encoding
t = poly2trellis(7,[171 133]);.
code = convenc(telbin,t); %
chan = rayleighchan(1e-8,4,[0 1e-7],[0 -14]);
delay = chan.ChannelFilterDelay;
M = 2;
pskSig = dpskmod(code,M);
fadedSig = filter(chan,pskSig);
rxsig=awgn(fadedSig,6);
rx = dpskdemod(rxsig,M)'
%decoding after receiving
qcode = quantiz(rx,[0.001,.1,.3,.5,.7,.9,.999]);
tblen = 48; delay= tblen; % Traceback length
decoded = vitdec(qcode,t,tblen,'cont','soft',3);
tx = telbin(2:end); rx1 = decoded(2:end);
tx_trunc = tx(1:end-delay); rx_trunc = rx1(delay+1:end);
[num,ber] = biterr(tx_trunc,rx_trunc)
a=zeros(1);
rx=decoded';
xanabin=reshape(rx,27540,8);
xanabin=xanabin+48;
kordoni=char(xanabin);
dek=bin2dec(kordoni);
leles=double(dek);
save rayconv.dat leles

```



```

%%Matlab code to simulate a Rician fading channel with AWGN and
%%convolutional coding on a Speex compressed file

clear
clc
fid=fopen('telmes16b','r+')
c=fread(fid,inf);
fclose(fid)
bin=dec2bin(c);
diplo=double(bin);
diplo2=diplo-48;
telbin=reshape(diplo2,1,(13773*8));
telbin=telbin';
t = poly2trellis(7,[171 133]);
%convolutional encoding
code = convenc(telbin,t);
chan = ricianchan(1e-8,10,30,[0 1e-7 1e-7],[0 -20 -20]);
delay=chan.ChannelFilterDelay;M = 2;
pskSig = dpskmod(code,M);
fadedSig = filter(chan,pskSig); rxsig=awgn(fadedSig,5);
rx = dpskdemod(rxsig,M);
%convolutional decoding
qcode = quantiz(rx,[0.001,.1,.3,.5,.7,.9,.999]);
tblen = 48; delay= tblen;
decoded = vitdec(qcode,t,tblen,'cont','soft',3);
tx = telbin(2:end); rx1 = decoded(2:end);
tx_trunc = tx(1:end-delay); rx_trunc = rx1(delay+1:end);
[num,ber] = biterr(tx_trunc,rx_trunc)
a=zeros(1);
rx=decoded';
xanabin=reshape(rx,13773,8);
xanabin=xanabin+48;
kordoni=char(xanabin);

```

```
dek=bin2dec(kordoni);
leles=double(dek);
save 5mion5.dat leles
```

```
%%Matlab code to simulate a fading channel with AWGN on an
%%uncompressed audio file

clc
clear
cwd = pwd;
cd(tempdir);
pack
cd(cwd)

%open file and measure sampling frequency
[y,fs,nbits]=wavread('wv5');
%sound(y,fs) to listen to file
y=y*(2^15);
megal=max(y);
mikr=min(y);
thetiko=y+(2^15);
megal=max(thetiko);
mikro=min(thetiko);
binar=dec2bin(thetiko);
diplo=double(binar);
diplo2=diplo-48;
siz1=size(diplo2);
telbin=reshape(diplo2,1,(11023*16));
telbin=telbin';

%insert Rayleigh channel
chan = rayleighchan(1e-8,90,[0 5e-8 1e-8],[0 -2 -3]);
delay = chan.ChannelFilterDelay;
M = 2;
pskSig = dpskmod(telbin,M);
```

```

fadedSig = filter(chan,pskSig);
%insert effects of channel
rx = dpskdemod(fadedSig,M);
tx = telbin(2:end); rx = rx(2:end);
tx_trunc = tx(1:end-delay); rx_trunc = rx(delay+1:end);
%measure BER
[num,ber] = biterr(tx_trunc,rx_trunc)
a=zeros(1);
rx=rx';
bx=[rx(1:end) a];
xanabin=reshape(bx,11023,16);
siz2=size(xanabin);
xanabin=xanabin+48;
kordoni=char(xanabin);
dek=bin2dec(kordoni);
leles=double(dek);
zzz=leles-(2^15);
mmm=zzz/(2^15);
%listen to the distorted file
sound(mmm,fs)

```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- [1] U. Black, *Voice Over IP*, Prentice Hall PTR, Upper Saddle River, 2000.
- [2] D. McMorries, "Investigation into the Effects of Voice and Data Convergence on a Marine Expeditionary Brigade TRI-TAC Digital Transmission Network," Master's thesis, Naval Postgraduate School, Monterey California, 2004.
- [3] H. Vegter, and D.W. Wallace, "Exploitation of Existing VoIP Technology for Department of the Navy Application," Master's thesis, Naval Postgraduate School, Monterey California, 2002.
- [4] T. Wallingford, *Switching to VoIP*, O'Reilly Media Inc., Sebastopole, 2005.
- [5] F. Hammer, P., Reichl, A. Raake, "The well-tempered conversation: interactivity, delay and perceptual VoIP quality," *IEEE International Conference on Communications, ICC 2005*, Volume 1, pp: 244 – 249, Seoul, May 2005.
- [6] T., Lutfullah, "Analysis of voice quality problems of Voice Over Internet Protocol (VoIP)," Master's thesis, Naval Postgraduate School, Monterey California, 2003.
- [7] Z., Li, L., Sun, Z., Qiao E., Ifeachor, "Perceived speech quality driven retransmission mechanism for wireless VoIP," *4th International Conference on 3G Mobile Communication Technologies 2003, 3G 2003*, pp: 395 – 399, London, June 2003.
- [8] M., Ergen, S., C., Ergen, P., Varaiya, "Throughput performance of a wireless VoIP model with packet aggregation in IEEE 802.11," *IEEE Wireless Communications and Networking Conference, . WCNC 2006*. Volume 4, pp: 2235 – 2239, Las Vegas, April 2006.
- [9] X., Holly, P., Zarrella, "Quality effects of wireless VoIP using security solutions" *IEEE Military Communications Conference, MILCOM 2004*. Volume 3, pp: 1352 – 1357, Monterey, November 2004.
- [10] D., Chairporn, "Performance evaluation of Voice over Internet Protocol," Master's thesis, Naval Postgraduate School, Monterey California, 2002.
- [11] H. Liu, and P. Mouchtaris, "Voice over IP Signaling: H323 and beyond," *IEEE Communications Magazine*, Vol. 38, Issue 10, pp. 142-148, October 2000.
- [12] M. Adda, A. Peart, and L. Rochfort, "Voice over IP on Wireless Networks," *Information and Communication Technologies, 2006, ICTTA '06*, Vol. 2, Damascus, 24-28 April 2006.
- [13] B. Goode, "Voice Over Internet Protocol," *Proceedings of the IEEE*, Vol. 90, Issue 9, pp. 1495-1517, September 2002.

- [14] J. Davidson, et. al., *Voice over IP fundamentals*, Cisco Press, Indianapolis, 2006.
- [15] D. Tamara, *Network+ Guide to Networks*, Thomson Course Technology, Boston, 2006.
- [16] D. Comer, *Internetworking with TCP/IP: Principles, Protocols, and Architecture, Volume 1*, Prentice Hall, Upper Saddle River, 2006.
- [17] Y. Zhang, "SIP based VoIP network and its internetworking with the PSTN," *Electronics and Communications Engineering Journal*, Vol. 14, Issue 6, pp. 273-282, December 2002.
- [18] R. Steinmetz, and K. Nahrstedt, *Multimedia: Computing, Communications and Applications*, Prentice Hall, Upper Saddle River, 1995.
- [19] K. Sayoud, *Introduction to data compression*, Morgan Kauffman, San Francisco, 2006.
- [20] A. Gersho, "Advances in Speech and Audio Compression," *Proceedings of the IEEE*, Vol. 82, Issue 6, pp. 900-918, June, 1994.
- [21] R. J. Deller, et al, *Discrete Time Processing of Speech Signals*, McMillan Press, London, 1993.
- [22] Cisco Systems Inc, "High Availability White Paper," January, 2007, http://www.cisco.com/en/US/products/hw/cable/ps2217/products_white_paper09186a00801af388.shtml.
- [23] H. Schulzrinne W. Jiang, "Assessment of VoIP Service Availability in the Current Internet," *Proceedings on the Passive and Active Measurement Workshop (PAM)*, pp. 151-157 La Jolla, California, April 2003.
- [24] Federal Communications Commission, "In the Matters of IP-enabled services E911 Requirements for IP-Enabled Service Providers," FCC 05-116, Washington D.C., June 3, 2005.
- [25] H. M. Chong, and H. S. Matthews, "Comparative Analysis of Traditional Telephone and Voice-over-Internet Protocol (VoIP) Systems," *Proceedings of the 2004 IEEE International Symposium on Electronics and the Environment*, pp: 106-111 , Scottsdale, May 2004.
- [26] J. Anderson, "Methods for Measuring Perceptual Speech Quality – Agilent Technologies White Paper," August, 2002 <http://cp.literature.agilent.com/litweb/pdf/5988-2352EN.pdf>.
- [27] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality," ITU Geneva, Switzerland, August 1996.

- [28] ITU-T Recommendation P.830, "Subjective Performance assessment of Telephone Band and Wideband Digital Codecs," ITU Geneva, Switzerland, February 1996.
- [29] D. E. Comer, *Computer Networks and Internets*, Prentice Hall, Upper Saddle River, 1997.
- [30] M. Skolnic, *Introduction to Radar Systems*, McGraw Hill, New York, 2001.
- [31] T. S. Rappaport, *Wireless Communications Principles and Practice*, Prentice Hall Communication Engineering and Emerging Technology Series, Upper Saddle River, 1996.
- [32] B. Sklar, *Digital Communications: Fundamentals and Applications*, Prentice Hall, Upper Saddle River, 2001.
- [33] W. Stallings, *Wireless Communications and Networks*, Prentice Hall, Upper Saddle River, 2005.
- [34] S. Haykin, *An Introduction to Analog and Digital Communications*, Willey and Sons, Indianapolis, 1989.
- [35] J. D. Beyda, *Data Communications from Basics to Broadband*, Prentice Hall, Upper Saddle River, 2000.
- [36] M. Roden, *Digital Communication Systems Design*, Prentice Hall, Upper Saddle River, 1988.
- [37] IEEE Standard 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer Specifications," ISO/IEC 802-11-1999, 1999.
- [38] S. McClure et al, *Hacking Exposed: Network Security Secrets and Solutions*, McGraw Hill, New York, 2005.
- [39] M. H. Manshaei, et al, "Performance Analysis of the IEEE 802.11 MAC and Physical Layer Protocol," Technical report, Institute National de Recherche en Informatique et en Automatique (INRIA), France, 2004.
- [40] T. Keller, and L. Hanzo "Adaptive Multicarrier Modulation: A Convenient Framework for Time-Frequency Processing in Wireless Communications," *IEEE Proceedings of the IEEE*, Vol. 88, Issue 5, pp.611-640, May 2000.
- [41] IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, "IEEE STD 802.16 – 2004," October 2004.
- [42] F. Ohrtman, *Wimax Handbook*, McGraw Hill, New York, 2005.

- [43] Speex Development Team, “Technical Documentation Manual”, 15 March 2007, <http://www.speex.org/docs/manual/speex-manual/node3.html>.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Professor Jeffrey B. Knorr
Department of Computer and Electrical Engineering
Naval Postgraduate School
Monterey, California
4. Professor Murali Tummala
Department of Computer and Electrical Engineering
Naval Postgraduate School
Monterey, California
5. Professor John McEachen
Department of Computer and Electrical Engineering
Naval Postgraduate School
Monterey, California
6. Embassy of Greece, Naval Attaché
Washington, DC
7. Tiantioukas Nikolaos
Zitsis 83
Ilion Attikis, Greece